PERGAMON

Vision Research 38 (1998) 1655–1682

Vision Research

Surface orientation from texture: ideal observers, generic observers and the information content of texture cues

David C. Knill *

Department of Psychology, University of Pennsylvania, 3815 Walnut Street, Philadelphia, PA 19104, USA

Received 23 October 1996; received in revised form 15 July 1997

Abstract

Perspective views of textured, planar surfaces provide a number of cues about the orientations of the surfaces. These include the information created by perspective scaling of texture elements (scaling), the information created by perspective foreshortening of texels (foreshortening) and, for textures composed of discrete elements, the information created by the effects of both scaling and foreshortening on the relative positions of texels (position). We derive a general form for ideal observers for each of these cues as they appear in images of spatially extended textures, (e.g. those composed of solid 2-D figures). As an application of the formulation, we derive a set of 'generic' observers which we show perform near optimally for images of a broad range of surface textures, without special prior knowledge about the statistics of the textures. Using simulations of ideal observers, we analyze the informational structure of texture cues, including a quantification of lower bounds on reliability for the three different cues, how cue reliability varies with slant angle and how it varies with field of view. We also quantify how strongly the reliability of the foreshortening cue depends on a prior assumption of isotropy. Finally, we extend the analysis to a naturalistic class of textures, showing that the information content of textures particularly suited to psychophysical investigation can be quantified, at least to a first-order approximation. The results provide an important computational foundation for psychophysical work on perceiving surface orientation from texture.

Keywords: Texture; Scaling; Slant; Foreshortening; Ideal observers; Slant-from-texture

1. Introduction

Gibson first pointed out fifty years ago that texture patterns can provide strong monocular cues to 3-D surface geometry [1]. More recent work has shown that in some contexts, these cues can be almost as salient for human observers as stereo [2,3] or motion cues [4]. While regular texture patterns such as checkerboards clearly give the most compelling perceptions of surface orientation and curvature, stochastic textures work well as 3-D cues also (see Fig. 1). Since the information provided by stochastic textures about surface geometry is inherently statistical, an important part of our understanding of texture information is an understanding of its theoretical limits. How reliably do different texture cues determine judgments of surface geometry? How does texture cue reliability vary as a function of underlying surface geometry? How does it vary as a function of viewing conditions, or as a function of surface texture properties? Answers to questions such as these provide an important computational back-drop for investigations of human perception of shape from texture.

In this study, I present an extended analysis of the statistical limits on the information provided by image texture patterns about planar surface orientation. It is a companion article to a psychophysical study [5] in which I investigate the limits of human ability to make judgments about planar surface slant from texture information. I derive ideal observers which optimally use texture information to estimate surface orientation and describe the results of simulations designed to measure the reliability of texture information in a variety of stimulus conditions. I also derive a set of generic estimators for different texture cues which we show to perform near-optimally for a wide range of texture ensembles.

* Corresponding author. Fax: +1 215 8987301; e-mail: knill@cattell.psych.upenn.edu.

0042-6989/98/$19.00
PII: S0042-6989(97)00324-6

The work presented here is an extension of previous ideal observer work [8,6,7,9] on the problem of shape from texture in a number of ways. First, I generalize the ideal observers from point element [8] and line-element [6,7,9] models of surface textures to textures composed of spatially extended texture elements. This follows the natural progression from zero to one to two dimensional texture descriptors. Second, I derive an ideal observer for perspective scaling information, a cue which has not, to date, been formally modeled in the ideal observer framework. Finally, I derive a set of generic ideal observers for different texture cues which I will show to be near-optimal estimators of surface orientation for a wide range of texture ensembles.

1.1. Aims of the study

Texture information is commonly decomposed into three quasi-independent cues to surface orientation and curvature: perspective scaling, foreshortening (or compression) and density (but see [11] for a shape-from texture method which does not, strictly speaking, rely on any decomposition). This decomposition will form the basic framework for my own work. I will derive individual ideal observers for surface orientation from each of the three cues and use these to analyze and compare the information content of the different cues.

The study is organized into six sections, including this introduction. In the second section, I briefly review the qualitative structure of texture information. In the third section, I derive a general form for the ideal observers for each of the three cues. The fourth section relies on the general ideal observer formulation to derive generic observers which I will later show to be near-optimal estimators for surface orientation from texture. The fifth section contains the bulk of the analysis, based on ideal and generic observer simulations. In this section, I will (1) compare texture informativeness about surface slant (orientation away from the line of sight) and tilt (orientation in the image plane), (2) analyze how changes in surface texture structure affect the reliability of texture information, including determining lower bounds on the reliability of texture cues given certain qualitative constraints on surface textures, (i.e. measuring the least information a cue can possibly provide about surface orientation given that textures are homogeneous and isotropic), (3) analyze the effects of field of view on the reliability of the information provided by the different texture cues and (4) analyze the importance of the isotropy constraint in determining the informativeness of texture foreshortening information. The final section of the study discusses the implications of the current results for computational models and human psychophysics.

2. Qualitative properties of texture information

Two distinctly different processes give rise to texture information in images. The first is the surface texture generating process, which reflects itself in statistical regularities in the ensemble of surface texture patterns it generates. The second is the image formation process; particularly, perspective projection, which maps surface textures to image textures. For in-surface textures, we can model image formation as mapping one 2-dimensional scalar field to another (ignoring spectral information)[12]. Assuming no marked shading effects on the scale at which texture properties are measured, we can describe image formation through the geometric distortions imposed by perspective. The form of the distortions are deterministically related to the shape and orientation of a surface relative to an observer, (e.g. scaling is inversely related to depth); thus, estimating the shape and orientation of a surface amounts to estimating the spatial pattern of texture distortion over an image. For image textures projected from stochastic surface textures, the problem is a classic problem in statistical estimation. Any model of texture information must therefore have two components: a deterministic model of perspective texture distortion and a stochastic model of surface textures. Combining the two, one can arrive at a model expressing the statistical relationship between image texture data and 3D surface geometry. In this section, I give a brief, tutorial overview of both components of the texture information model we use.

2.1. Perspective projection

Gibson conceived of texture information as spatial gradients of various types; density, size, etc. which arise as a result of perspective projection. While this matches our intuitions about texture information, it is only exact for regular texture patterns. A more meaningful way to interpret texture gradients is to treat the gradients which appear in images of regular textures as illustrative of the underlying geometry of perspective projection. In this view, while it is not appropriate to characterize texture information per se in the form of gradients, it is appropriate and very useful to characterize perspective projection in this way, using stereotypical texture patterns, such as arrays of circles, to illustrate the distorting effects of perspective.

Locally, perspective projection distorts a texture pattern in two distinct ways: by scaling the texture and by distorting its shape. The 'size' of a local texture patch is scaled by an amount inversely proportional to the distance of the patch from the nodal point of the eye. Similarly, the shape of the texture patch is foreshortened by an amount proportional to the cosine of the slant of the surface away from the local line of sight. Both of these effects are first-order approximations to

the local perspective distortion. They vary with spatial position in a predictable way as a function of surface geometry. For slanted, planar surfaces, distance from the viewer changes as a simple function of position in the image, as does local slant, since the angle of the local line of sight changes across an image. These two factors give rise to spatial variations in the local distortion of a texture pattern.

2.2. Statistical structure of surface textures

The scaling and foreshortening components of the local texture map (and their 'gradients') are not directly available to an observer, who only sees the effects of these distortions. Without exact a-priori knowledge of the spatial structure of a surface texture, the form of local distortion and hence the local 3D shape, cannot be directly inferred from image data. Since most surface textures are stochastic, the best any observer can hope for is knowledge of the statistical properties of the ensemble from which a surface texture is drawn. Such knowledge supports statistical best guesses about the local distortion. The two general statistical constraints which can make texture informative are homogeneity and isotropy.

2.2.1. Homogeneity

Homogeneity, formally defined, means that a stochastic surface texture process is spatially stationary: that the statistical relationships between points on the surface depend only on their relative positions, not on their absolute position in some global reference frame. For most doubly-curved surfaces, homogeneity is very difficult to define unambiguously[1]. For planar surfaces, however, with which we are presently concerned, the definition is straightforward: the statistics of homogeneous, planar textures are invariant to translations in the plane of the surface. If a planar surface has a homogeneous surface texture, then perspective projection induces particular spatial variations in the local statistical structure of an image texture which are related to the slant and tilt of the surface. An observer can, therefore, use variations in the sample statistics of an image texture to make probabilistic inferences about surface slant and tilt.

2.2.2. Isotropy

Isotropy refers to a lack of directional bias in the statistics of surface textures. How isotropy is applied in computational work varies from model to model, but one can formalize a general meaning for planar textures in terms of the marginal probability laws relating texture properties at different points in the plane (see, for example, [14]). In particular, a strictly isotropic 2-dimensional texture process is one whose marginal probability laws are invariant to rotations of the coordinate frame in which the process is defined. Isotropy is the statistical analogue of exact knowledge of local texture geometry; it allows an observer to infer the perspective texture distortion locally, in this case, by using the average amount of directional bias in a local patch of image texture to infer the amount of foreshortening and hence the surface orientation.

All researchers who deal with the problem of perceiving surface geometry from texture consider homogeneity a minimal assumption necessary to make image textures informative. Prior knowledge of isotropy, when appropriate, adds considerably to the informativeness of textures. Not all textures are homogeneous and only some of these are isotropic; however, one may well assume that a large class of textures are homogeneous and some sub-class of these are isotropic, so that both the homogeneity and isotropy constraints would prove useful to an observer, as long as they could test their validity before applying them [9] (as is true of most natural constraints: rigidity, symmetry, etc.).

3. Ideal observers for surface orientation from texture

I define an ideal observer for surface orientation from texture to be the maximum likelihood estimator of surface orientation from a set of image texture measurements. My approach to the problem of deriving ideal observers is to begin with a completely general formulation of the likelihood function and gradually refine it by defining first, a representational framework for surface and image textures and second, a model of perspective projection. This will lead to a more explicit model of the likelihood function, which is incomplete only in that it requires a specification of a prior stochastic model of surface textures. I then show that, given some independence assumptions on the surface texture model, the likelihood function for surface orientation from texture can be decomposed into the product of three marginal likelihood functions, one each for the three texture cues defined in the introduction. The different marginal likelihood functions serve as the basis for three distinct ideal observers; again, one for each cue. In the section following this one, I will use the likelihood functions to derive a set of generic observers, which are designed to be near-optimal for a large set of surface texture models.

The section is organized into four parts: a review of previous ideal observer work, a derivation of the general form for likelihood functions for surface orientation from texture, a specification of the major

[1] Defining what it means to say that texture statistics are the same everywhere on a surface is problematic for anything but constant-Gaussian curvature surfaces [11] and even for those is limited when the Gaussian curvature is other than zero [13].

components of the general likelihood function, which I will use as the basis for our analysis of texture information and a decomposition of the likelihood function into component likelihood functions for each of the three texture cues.

3.1. Previous ideal observer work

A number of authors have derived ideal estimators of surface orientation from texture based on prior stochastic models of surface textures. These formulations have been limited to textures made up of independent point elements or independent, oriented line elements [6–9]. The ideal observers were defined for surface texture ensembles characterized by default prior models: independent and uniform distributions for both position and orientation. Given the qualitative constraints that surface textures are homogeneous and isotropic, these priors represent maximum entropy models of surface texture processes; that is, they are the least constrained prior models one can define (reflecting a principle of least commitment to prior assumptions). The combination of point element and line element models forms the most currently developed statistical ideal observer model for shape from texture [10]. The model accurately characterizes the information provided by the density and foreshortening cues in images of textures consisting of collections of independent and uniformly positioned and oriented line elements on a surface.

Ideal observer models are distinguished by two features. First, they are maximum likelihood estimators (MLEs) for surface shape and orientation for particular classes of textures; thus, they are optimal estimators (in the minimal variance sense[2]) for the classes of textures for which they were derived. Second, they typically deal only with the inference component of the problem of estimating shape from texture. They do not account for the actual measurement of texture primitives from real images; thus, they provide baseline measures of texture information assuming the existence of some form of ideal measurement system. Most existing models of shape from texture do not fit the strict optimality criteria of an ideal observer. This is in large part due to the fact that many models integrate image measurement with shape estimation [16–19,11], a context in which it is difficult to define prior stochastic models of surface textures and derive maximum likelihood estimators from first principles. A review of these models is beyond the scope of this study.

[2] MLEs are only (provably) minimal variance estimators in the limit as the sample size goes to infinity; however, textures with large numbers of elements approximate this optimality condition.

3.2. The likelihood function for surface orientation from texture

I define the ideal observer for estimating a set of surface geometry parameters, $\vec{S}$, from a set of image texture measurements, $\vec{T}_{\mathrm{I}}$, to be the maximum likelihood estimator of $\vec{S}$ given the image measurements[3]. For my purposes, $\vec{S}$ is the slant and tilt of a planar surface relative to some global coordinate frame, though, in general, it could include other parameters to characterize the shapes of curved surfaces. We will represent surface orientation as the slant-tilt pair, (σ, τ) (slant is the absolute angle of the surface away from the fronto-parallel plane and tilt is the angle it makes in the fronto-parallel image plane). The maximum likelihood estimator of surface orientation selects as its estimate that value of the pair (σ, τ) which maximizes the likelihood function, $p(\vec{T}_{\mathrm{I}}|\sigma, \tau)$.

In order to derive a general form for the likelihood function, we assume that the projective mapping from surface texture parameters to measured image texture parameters is one-to-one, differentiable and invertible and write it as

$$\vec{T}_{\mathrm{I}} = \pi(\vec{T}_{\mathrm{S}}; \sigma, \tau) \tag{1}$$

where $\vec{T}_{\mathrm{S}}$ is a representation of the surface texture. We further assume noise-free measurements of image texture parameters (so that the uncertainty in texture information derives entirely from the random structure of surface textures). This allows us to write a likelihood function for $\vec{T}_{\mathrm{I}}$ conditioned on both surface slant and tilt and on the surface texture, $\vec{T}_{\mathrm{S}}$, as a delta function

$$P(\vec{T}_{\mathrm{I}}|\sigma, \tau, \vec{T}_S) = \delta(\vec{T}_{\mathrm{I}} - \pi(\vec{T}_{\mathrm{S}}; \sigma, \tau)) \tag{2}$$

We could use this likelihood function to estimate both surface orientation and the surface texture parameters, $\vec{T}_S$. This would amount to finding a solution which satisfies Eq. (1); an underdetermined problem, since for every surface orientation, there exists some set of texture parameters which will make the argument of the delta function go to zero (by assumption of the invertibility of π).

We are only interested in estimating the surface orientation and do not care about the surface texture, per se. One can, therefore, integrate out the surface texture variables to arrive at a likelihood function for surface orientation only,

$$p(\vec{T}_{\mathrm{I}}|\sigma, \tau) = \int_{\psi} p(\vec{T}_{\mathrm{I}}|\sigma, \tau, \vec{T}_{\mathrm{S}})\, p_{T_{\mathrm{S}}}(\vec{T}_{\mathrm{S}})\, \mathrm{d}\vec{T}_S \tag{3}$$

[3] The maximum liklihood estimator of a variable $\vec{S}$ approximates the maximum a-posteriori (MAP) estimator (which takes into account the prior distribution on $\vec{S}$) when the spread of the prior distribution is much greater than the support of the liklihood function, a condition which is typically met when large numbers of image measurements are available as data.

$$p(\vec{T}_I|\sigma,\tau)=\int_{\psi}\delta(\vec{T}_I-\pi(\vec{T}_S;(\sigma,\tau))\,p_{T_S}(\vec{T}_S)\,d\vec{T}_S \quad (4)$$

where $p_{T_S}(TS)$ is the prior distribution of $\vec{T}_S$ and ψ is the space of possible values of $\vec{T}_S$. The result gives for the likelihood function,

$$p(\vec{T}_I,|\sigma,\tau)$$
$$=p_{T_S}(\vec{T}_S=\pi^{-1}(\vec{T};(\sigma,\tau)))\,\det[J_T(\pi^{-1}(\vec{T};\sigma,\tau))] \quad (5)$$

where $\pi^{-1}()$ is the inverse of the projection function and $J_T(\pi^{-1}())$ is the Jacobian of $\pi^{-1}()$ with respect to the image texture variables; that is, the matrix containing the first partial derivatives of $\pi^{-1}()$ with respect to the image texture variables. $p_{T_S}()$ is the prior probability on the surface texture parameters.

The two terms in Eq. (5) reflect the relative contributions of the prior stochastic model of surface textures and of the projection model to the likelihood function. The prior term gives preferential weighting to surface orientations for which the back-projected surface texture best fits the stochastic model of surface textures used to generate the stimuli. The Jacobian term rewards surface orientations for which image texture measurements are least sensitive to changes in the surface texture parameters. This reflects a preference for orientations which explain a broader range of possible values for the surface texture parameters [20]. Specifying an ideal observer for an ensemble of surface textures requires a model of both components of the likelihood function.

3.3. An explicit model of texture information

3.3.1. A moment-based texture representation

The first step in fleshing out a specific ideal observer model for surface orientation from texture is to define the primitive elements of which surface and image textures are composed. Most psychophysical work on the perception of surface layout from texture has relied on textures composed of discrete elements (texels). Such textures are particularly easy to synthesize for use in experiments and to analyze using ideal observer analysis; thus, we will follow in this tradition. In general, texels can be quite complex in their shape, as is true for the polygonal tiles making up the texture shown in Fig. 1. Thus, we require a set of measurements on both surface and image texels which provide a description of texel shape which on the one hand, is compact, and on the other hand, can be derived for arbitrary texels (polygons, closed, smooth curves, etc.).

In choosing a representational framework for characterizing surface and image texels, I was driven by two general constraints.

- The framework must support a complete model of projection. Stating this formally, we must be able to model the projection of a surface texel into the image as a bijective mapping between the representation of a surface texel, $\vec{T}_S$ and the representation of the image texel to which it projects, $\vec{T}_I$; that is, the function, $\vec{T}_I=f(\vec{T}_S;\sigma,\tau)$ must be both on-to-one and invertible (for a given surface orientation, (σ,τ)). An example of a texel representation which violates this constraint is one which specifies only the dominant orientation of a texel, since the orientation of a texel in the image is determined not only by the orientation of the corresponding texel on the surface, but also by its shape (except for infinitely thin line elements).
- The parameters included in the representation of image texels should be qualitatively similar to those which could be plausibly derived from local measurements of continuous textures. This will allow the derived ideal observers to be applied to the outputs of a putative texture measurement system.

The simplest parameterization of texel shape which satisfies both constraints derives from the second-order spatial moments of a texel. Moment-based formulations, because of the ease with which they can be estimated from image data, have recently been popular as a basis for computational models of shape from texture [9,19,21] and can be implemented using biologically plausible filter mechanisms [17]. I use an area-based formulation of the second-order moments[4], which is characterized by the moment tensor,

$$M=\frac{1}{A}\begin{pmatrix}\int\int_R(x-x_0)^2\,dA & \int\int_R(x-x_0)(y-y_0)\,dA\\ \int\int_R(x-x_0)(y-y_0)\,dA & \int\int_R(y-y_0)^2\,dA\end{pmatrix} \quad (6)$$

where R is the region bounded by a texel, A is the area of the texel, given by

$$A=\int\int_R dA \quad (7)$$

and (x_0,y_0) is the center of mass of the texel, given by

$$(x_0,y_0)=\tfrac{1}{A}\left(\int\int_R dA,\int\int_R y\,dA\right) \quad (8)$$

In Appendix B, we show that the second-order moments of an image texel, M_I, can indeed be expressed as a bijective function of the second-order moments of the corresponding surface texel, M_S, independent of the detailed shape characteristics of the texels (see Fig. 2).

[4] A boundary-based formulation of second-order moments does not support a general model of projection, due to the arc-length foreshortening effects of projection. These effects will reflect themselves in different ways depending on the detailed shape of a surface texel.

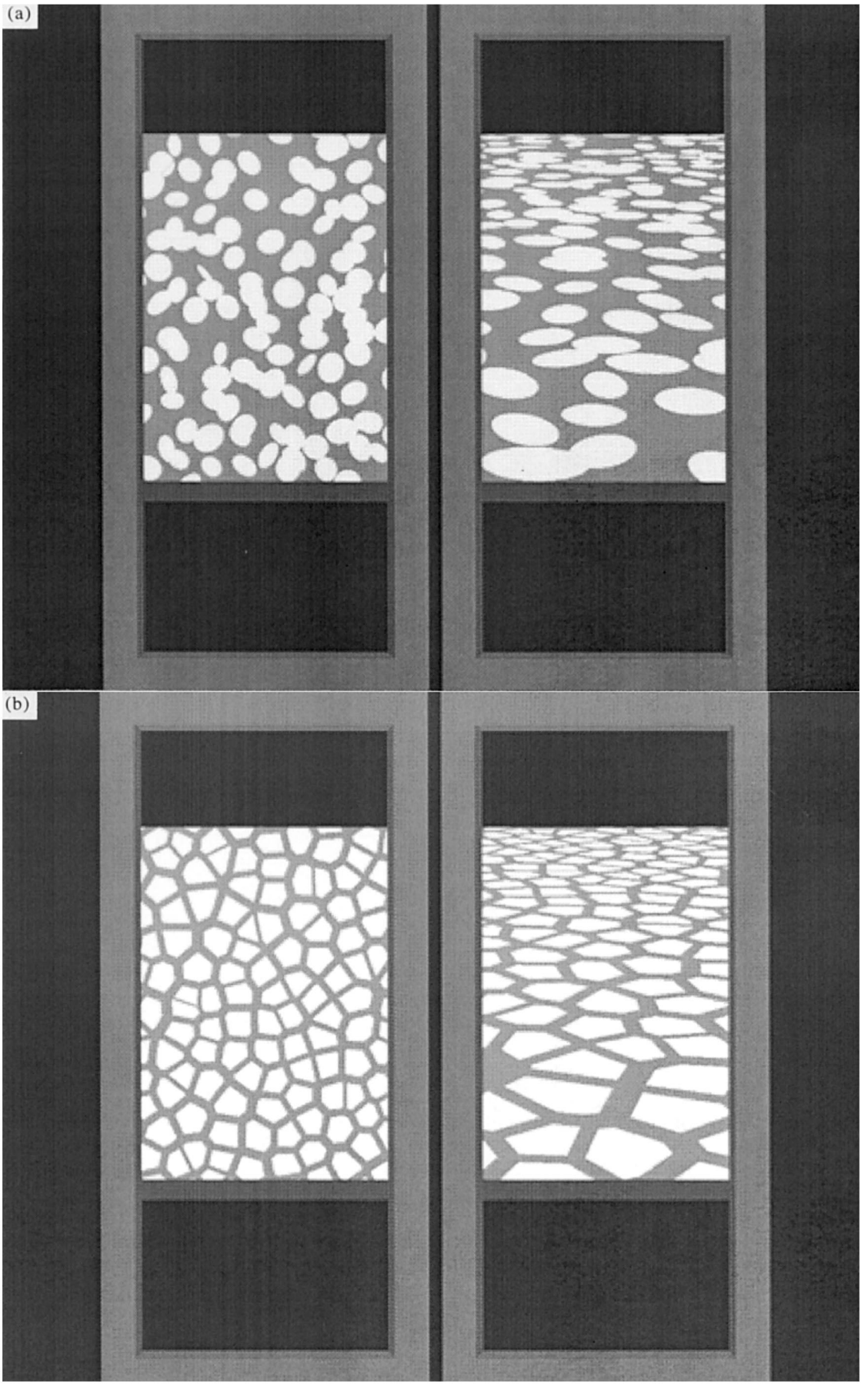

Fig. 1. Examples of texture patterns which induce strong percepts of surfaces slanted in depth. In each of the two figures, one surface is fronto-parallel and the other is slanted away from the fronto-parallel at 73°. The images in (a) were created by randomly drawing ellipses on a surface. The images in (b) were generated from the Voronoi diagram derived from a random lattice of points (see Section 5.2.3 for a description of Voronoi diagrams).

3.3.2. The projection model

For in-surface textures, a natural way to model perspective projection is as a map between points on a surface and points in the image, $\pi:\Sigma \rightarrow \Omega$, where Σ is the surface and Ω is the image. For planar surfaces, we can easily characterize the projective map as a function, $f()$, which maps a 2-D position vector on the surface (in a coordinate system defined on the plane) to its corresponding 2D position in an image (see Appendix A for a derivation of $f()$ for planar perspective),

$$\vec{X} = f(u, v) = \begin{pmatrix} x(u, v) \\ y(u, v) \end{pmatrix} \tag{9}$$

where $\vec{U} = (u, v)^T$ represents position on the surface and $\vec{X} = (x, y)^T$ represents position in the image.

We can use this simple formulation of texture projection to derive a first-order approximation to the mapping of a surface texel to an image texel (see Fig. 2). Define a surface texel to be a bounded region on the surface, $R_S \in \Sigma$, and an image texel to be a bounded region in the image, $R_I \in \Omega$, which results from the projection of R_S; that is, $R_I = \pi(R_S)$. Since I am representing image texels by their second-order spatial moments, I am particularly interested in specifying the mapping between the second-order moments of a surface texel and the second-order moments of its projection. In Appendix B, I show that a first-order approximation of this mapping (exact in the limit for infinitely small texels), is given by

$$M_I = P^{-1} M_S (P^{-1})^T \tag{10}$$

where M_I is the second-order moment tensor for the image texel, M_S is the second-order moment tensor for the surface texel and P is the differential of the projective map computed at the center of mass of R_S. P is given by the Jacobian of $f()$,

$$P = \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix} \tag{11}$$

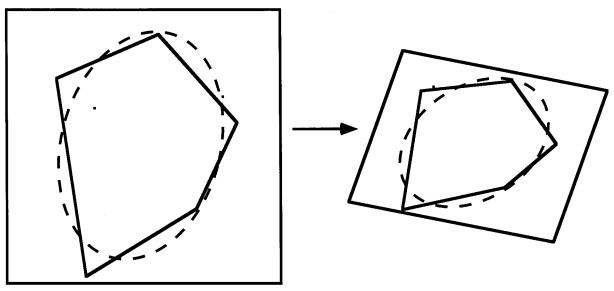

Fig. 2. Representing the shapes of irregular texels by their second-order spatial moments amounts to fitting them with ellipses. A first-order approximation to the mapping from surface texels to image texels is an affine transformation between a fitted ellipse on the surface and a fitted ellipse in the image.

Appendix A derives P for the case of planar perspective. I write the result here for a surface tilted in the vertical direction (like a ground plane), expressed as a function of the center of mass of an image texel, (x, y),

$$P = \left(1 - \frac{y}{d}\tan\sigma\right) \begin{pmatrix} 1, & -\frac{x}{d}\sin\sigma \\ 0, & \cos\sigma - \frac{y}{d}\sin\sigma \end{pmatrix} \tag{12}$$

where σ is the slant of the surface away from the fronto-parallel.

The scalar term captures the effect of perspective scaling. It is a decreasing function of height in the image (texels shrink with height in the image). The non-constant terms in the matrix represent the distortion of texel shape induced by perspective foreshortening. Texels are compressed in the vertical direction by an amount determined by the height in the image, but are also skewed (for $x \neq O$) by an amount determined by horizontal position, as reflected in the off-diagonal term in P. Eq. (12) can be easily generalized to surfaces with variable tilt by an appropriate rotation of P,

$$P = \left(1 - \frac{y}{d}\tan\sigma\right) \begin{pmatrix} \cos\tau, & \sin\tau \\ \sin\tau, & \cos\tau \end{pmatrix} \times \begin{pmatrix} 1, & -\frac{x}{d}\sin\sigma \\ 0, & \cos\sigma - \frac{y}{d}\sin\sigma \end{pmatrix} \tag{13}$$

where τ is the tilt of the surface.

Eq. (10) specifies the projection model which maps the moment representation of a surface texel to the moment representation of the corresponding image texel. As shown in Appendix B, this mapping is independent (in the infinitesimal limit) of the exact shape of a texel. Perhaps the easiest way to visualize the projection model is to consider how it distorts a unit circle on the surface. If the circle is given by the implicit quadratic

$$(\vec{U} - \vec{U}_0)^T (\vec{U} - \vec{U}_0) = 1, \tag{14}$$

then its image is the ellipse defined by the implicit quadratic equation

$$(\vec{X} - \vec{X}_0)^T (P^T)^{-1} P (\vec{X} - \vec{X}_0) = 1 \tag{15}$$

where $(P^T)^{-1}P$ is the quadratic form of the ellipse and $\vec{X}_0$ is the projection in the image of $\vec{U}_0$.

3.3.3. Decomposing the likelihood function

In this section, I will formalize the definitions of the three texture cues by decomposing the likelihood function for surface orientation from texture into the product of three marginal likelihood functions, one each for the scaling, foreshortening and position cues. The basic

observation behind the decomposition is that representing texels by their second-order spatial moments amounts to fitting the texels with ellipses. I can therefore reformulate the representation of a texel in terms of the length, aspect ratio and orientation of its best-fitting ellipse, as determined by the eigen-vectors of M_S (for surface texels) and M_I (for their projections in the image). This suggests a parameterization of surface textures as sets of five-tuples, $\vec{T}_S = \{(u_1, v_1, l_1, \theta_1, a_1), \ldots, (u_n, v_n, l_n, \theta_n, a_n)\}$, where (u_i, v_i) specifies the position of texel i, l_i its length, θ_i its orientation and a_i its aspect ratio. Similarly, for image textures, we can write $\vec{T}_I = \{(x_1, y_1, \lambda_1, \phi_1, \alpha_1), \ldots, (x_n, y_n, \lambda_n, \phi_n, \alpha_n)\}$.

In order to decompose the likelihood function, $p(\vec{T}_I|\sigma, \tau)$, into the product of a set of marginal likelihood functions, we require a number of independence assumptions. Most of these follow naturally from the assumptions of surface texture homogeneity and isotropy. By the definition of homogeneity, the texel shape and size parameters must be independent of position. Furthermore, if we assume isotropy, texel orientation must be independent of length and aspect ratio (since isotropy implies that the probability laws characterizing a texture are independent of the orientation of the coordinate system in which it is defined). Finally, I will assume that texel length (a scale parameter) is independent of texel aspect ratio (a scale-invariant shape parameter).

The likelihood function is given by the density function,

$$p(\vec{T}_I|\sigma, \tau) = p(\{x_i, y_i, \lambda_i, \phi_i, \alpha_i\}|\sigma, \tau) \tag{16}$$

The independence assumptions support decomposing the likelihood function into three marginal likelihood functions, one for texel length, one for texel shape (aspect ratio and orientation) and one for texel position,

$$p(\vec{T}|\sigma, \tau) = p(|x_i, y_i, \lambda_i, \phi_i, \alpha_i\}|\sigma, \tau) \tag{17}$$

$$p(\vec{T}|\sigma, \tau) = p(\{\lambda_i\}|\sigma, \tau; \{x_i, y_i, \phi_i, \alpha_i\}) \times p(\{\phi_i, \alpha_i\}|\sigma, \tau; \{x_i, y_i\})\, p(\{x_i, y_i\}|\sigma, \tau) \tag{18}$$

The marginal distributions characterize the information about surface orientation provided by the scaling, shape and position, respectively. The decomposition allows us to model individual ideal observers for surface orientation from each of three texture cues (each based on a different marginal likelihood function).

3.3.3.1. Size likelihood. For a given surface orientation, we can back-project an image texel onto the surface, giving a surface texel with length;

$$l_i = f(\lambda_i; x_i, y_i, \phi_i, \alpha_i, \sigma, \tau) \tag{19}$$

where $f()$ is simply one over the smallest eigen-value of the back-projected quadratic form. Using the general form for the likelihood function in Eq. (5), we can write the likelihood for texel length as

$$p(\{\lambda_i\}|\sigma, \tau; \{x_i, y_i, \phi_i, \alpha_i\}) = p(\{l_i = f(\lambda_i; x_i, y_i, \phi_i, \alpha_i, \sigma, \tau)\}) \prod_{i=1}^{n} \frac{\partial f(\lambda_1; \ldots)}{\partial \lambda_i} \tag{20}$$

For a given image texel position and shape, the back-projected surface texel length is proportional to the length of the image texel; therefore, we can write the derivative as the ratio, l_i/λ_i, giving for the likelihood function,

$$p(\{\lambda_i\}|\sigma, \tau; \{x_i, y_i, \phi_i, \alpha_i\}) = p(\{l_i = f(\lambda_i; x_i, y_i, \phi_i, \alpha_i, \sigma, \tau)\}) \prod_{i=1}^{n} \frac{l_i}{\lambda_i} \tag{21}$$

3.3.3.2. Shape likelihood. The back-projection of an image texel onto a surface with a given orientation results in a surface texture whose shape can be expressed as a function of the shape and position of the image texel and the slant and tilt of the surface,

$$(\theta_1, a_i) = g(\phi_i, \alpha_i; x_i, y_i, \sigma, \tau) \tag{22}$$

$g(.)$ is a vector valued function giving the orientation and aspect ratio of the surface texel (derived from the quadratic form for the back-projected texel). The likelihood function for texel shape is given by

$$p(\{\phi_i, \alpha_i\}|\sigma, \tau; \{x_i, y_i\}) = \prod_i^n p((\theta_i, a_i) = g(\phi_i, \alpha_i; x_i, y_i, \sigma, \tau)) \times \det[J_{\phi_i \alpha_i}(g(\phi_i, \alpha_i; x_i, y_i, \sigma, \tau))] \tag{23}$$

$$p(\{\phi_i, \alpha_i\}|\sigma, \tau; (\{x_i, y_i\}) = \prod_i^n \left(p((\theta_i, a_i) = g(\phi_i, \alpha_i; x_i, y_i, \sigma)) \left(\frac{\partial \theta_i}{\partial \phi_i} \frac{\partial a_i}{\partial \alpha_i} - \frac{\partial \theta_i}{\partial \alpha_i} \frac{\partial a_i}{\partial \phi_i} \right) \right) \tag{24}$$

Note that this likelihood function cannot be decomposed further into likelihood functions for texel orientation and aspect ratio, because the two texel properties are confounded in the projective mapping to images (the Jacobian is not diagonal).

3.3.3.3. Position likelihood. The back-projection of image texel positions onto a surface is given by the inverse of the projection function, π,

$$(u_i, v_i) = \pi^{-1}(x_i, y_i; \sigma, \tau) \tag{25}$$

The likelihood function for position information is then given by

$$p(\{x_i, y_i\}|\sigma, \tau) = p(\{(u_i, v_i) = \pi^{-1}(x_i, y_i; \sigma, \tau)\}) \times \prod_i^n J(\pi^{-1}(x_i, y_i; \sigma, \tau)) \tag{26}$$

$$p(\{x_i, y_i\}|\sigma, \tau) = p(\{(u_i, v_i) = \pi^{-1}(x_i, y_i; \sigma, \tau)\}) \prod_i^n \frac{\partial u_i}{\partial x_i} \frac{\partial v_i}{\partial y_i} \tag{27}$$

A simple way to visualize the Jacobian is as the back-projected area of a unit square in the image plane.

3.3.4. A note on notation

The ideal observer formulation I have just described decomposes texture information into cues characterized by one or another feature of image texels. In presenting this work to others, however, I have found that people find the image based terms for the texture cues more confusing than the standard terms. I will therefore use the terms scaling to refer to the information provided by texel lengths and foreshortening to refer to the information provided by texel shapes and orientations. Because the texel position ideal is only loosely related to density information, I prefer referring to texel position information as such, rather than as density information.

4. Generic observers

As a first application of the general form of the ideal observers described above, I will derive a set of generic estimators for surface orientation from each of the three texture cues described above. The idea is to derive estimators which assume as little prior knowledge as possible about the prior surface texture model.

4.1. Scaling

The only absolute constraint one can impose on texel lengths is that they are positive; thus the range of texel length is the half-line, $l > 0$. Much like Blake et al. did for density and line element orientation, I will derive a generic observer for scaling information by assuming a maximum entropy prior for texel length, which, for positive, real-valued variables, is the exponential distribution ([22]). Thus, we have for the prior on l,

$$p(l) = \tfrac{1}{k} e^{-l/k} \tag{28}$$

where k is a scale constant. Both the mean and standard deviation of this distribution are equal to k. The likelihood function for texel length using the exponential prior is given by the equation

$$p(\{\lambda_i\}|\sigma, \tau; \{x_i, y_i, \phi_i, \alpha_i\}) = p(\{\, l_i = f(\lambda_i; x_i, y_i, \phi_i, \alpha_i, \sigma, \tau)\}) \prod_{i=1}^n \frac{l_i}{\lambda_i} \tag{29}$$

$$p(\{\lambda_i\}|\sigma, \tau; \{x_i, y_i, \phi_i, \alpha_i\}) = \prod_{i=1}^n \tfrac{1}{k} \exp[f(\lambda_i; x_i, y_i, \phi_i, \alpha_i, \sigma, \tau)\})/k] \prod_{i=1}^n \tfrac{l_i}{\lambda_i} \tag{30}$$

recalling that $f(\lambda_i; x_i, y_i, \phi_i, \alpha_i, \sigma, \tau)$ is the back-projection function mapping image texel length to surface texel length. Eq. (30) has a free parameter in the prior term, the scale constant, k. Since we do not want to assume knowledge of the scale of the texture, we should integrate the scale constant out to arrive at a truly generic prior term. In order to do this, we need to assume some prior density function for k. The natural default prior for a scale variable is $1/k$.

Integrating over k, we obtain for the prior term in the likelihood function

$$p(\{l_i = f(\lambda_i; x_i, y_i, \phi_i, \alpha_i, \sigma, \tau)\}) = \frac{n!}{[\sum_{i=1}^n f(\lambda_i; x_i, y_i, \phi_i, \alpha_i, \sigma, \tau)]^n} \tag{31}$$

We can express the denominator in terms of the average back-projected texel length, $m_l(\sigma, \tau)$, giving for the prior,

$$p(\{l_i = f(\lambda_i; x_i, y_i, \alpha_i, \sigma, \tau)\}) = \frac{n!}{n^n} \frac{1}{m_l(\sigma, \tau)^n} \tag{32}$$

and for the generic likelihood function for texel size (after some algebraic manipulation),

$$p(\{\lambda_i\}|\sigma, \tau; \{x_i, y_i, \phi_i, \alpha_i\}) = \frac{n!}{n^n} \prod_{i=1}^n \frac{l_i}{m_l(\sigma, \tau)\, \lambda_i} \tag{33}$$

The appearance of the mean back-projected length in Eq. (33) highlights the fact that the generic prior acts to normalize the likelihood function, effectively rendering the estimator scaleinvariant. The generic likelihood function penalizes interpretations which give larger variances, relative to the mean, of back-projected texel lengths. This captures the heuristic intuition that perspective projection tends to increase the variance of texel lengths, so that the best interpretation of surface orientation should be the one corresponding to a minimal variance for back-projected texel lengths.

4.2. Foreshortening

In order to derive a generic estimator for surface orientation from foreshortening information, we must include both the orientations and aspect ratios of texels in the formulation. The isotropy constraint fully determines the prior distribution of texel orientations (a uniform distribution). The distribution of aspect ratios, however, cannot be known, a-priori. In order to derive a generic estimator for the foreshortening cue which makes full use of the information provided by texel aspect ratios, I will define a general family of prior distributions for aspect ratios and integrate the resulting likelihood function over the free parameters which define the prior distributions within the family, much as I did for texel length information. The design strategy

is motivated by two general observations: first, the probability density function for aspect ratios should go smoothly to zero at either extreme. At one extreme, for aspect ratios equal to zero, this follows from the vanishingly small probability of having physical textures with zero width. At the other extreme, for aspect ratios equal to one, the constraint derives from the fact that the Jacobian for texel shape and orientation has a singularity at one. Allowing a non-zero probability density at one would lead to an ill-conditioned estimator.

The general constraints just described have led us to model the prior distribution of aspect ratios in a transformed domain in which the range of aspect ratios of [0, 1] maps to the range $(-\infty, \infty)$. The simple transform

$$x = \log_e\left(\frac{a}{1-a}\right) \tag{34}$$

achieves this. I will assume that the prior on x is Gaussian with unknown mean and standard deviation. The resulting family of priors for aspect ratio covers a broad range; from roughly uniform to approximately Gaussian, but has the property that all distributions in the family go smoothly to zero as the aspect ratio approaches zero or one. This gives a prior of the form

$$p(a; \mu_x, \sigma_x) = \tfrac{1}{\sqrt{2\pi}\sigma_x} \exp\left[-\frac{(x(a)-\mu_x)^2}{2\sigma_x^2}\right]\frac{\partial x}{\partial a} \tag{35}$$

$$p(a; \mu_x, \sigma_x) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left[-\frac{(x(a)-\mu_x)^2}{2\sigma_x^2}\right]\left(\frac{1}{a-a^2}\right) \tag{36}$$

where $x(\alpha)$ is given by Eq. (34) and μ_x and σ_x are the mean and standard deviation of the transformed aspect ratio.

The likelihood function for foreshortening information can then be written as

$$p(\{\phi_i, \alpha_i\}|\sigma, \tau; \{x_i, y_i\}) = \prod_{i=1}^{n} \frac{1}{\pi} p\left(a_i = g(\phi_i, \alpha_i; x_i, y_i, \sigma, \tau)\left(\frac{\partial\theta_i}{\partial\phi_i}\frac{\partial a_i}{\partial\alpha_i} - \frac{\partial\theta_i}{\partial\alpha_i}\frac{\partial a_i}{\partial\phi_i}\right)\right) \tag{37}$$

$$p(\{\phi_i, \alpha_i\}|\sigma, \tau; \{x_i, y_i\}) = k \prod_{i=1}^{n} \frac{1}{\sigma_x} \exp\left[-\frac{(x(a_i)-\mu_x)^2}{2\sigma_x^2}\right]\left(\frac{1}{a_i-a_i^2}\right) \times \left(\frac{\partial\theta_i}{\partial\phi_i}\frac{\partial a_i}{\partial\alpha_i} - \frac{\partial\theta_i}{\partial\alpha_i}\frac{\partial a_i}{\partial\phi_i}\right) \tag{38}$$

where I have collected constants into a single term, k.

One can integrate out the free parameters in the prior term to arrive at a likelihood function for the foreshortening cue which depends on no prior knowledge of the form of the prior distribution. To do this, assume a uniform prior on μ_x and a $1/\sigma_x$ prior on σ_x. The latter derives from the observation that values of σ_x much greater than one lead to strongly bi-modal distributions of aspect ratios, an unlikely occurrence in the world. In point of fact, The exact form of the prior has little effect on the resulting likelihood function for large numbers of texels—assuming a uniform prior on σ_x changes the likelihood function only slightly and leads to an estimator whose performance is numerically indistinguishable from the estimator derived here. Carrying through the integration gives

$$p(\{\phi_i, \alpha_i\}|\sigma, \tau; \{x_i, y_i\}) = \prod_{i=1}^{n} \left(\frac{1}{s_x(\sigma, \tau)}\right)\left(\frac{1}{a_i-a_i^2}\right)\left(\frac{\partial\theta_i}{\partial\phi_i}\frac{\partial a_i}{\partial\alpha_i} - \frac{\partial\theta_i}{\partial\alpha_i}\frac{\partial a_i}{\partial\phi_i}\right) \tag{39}$$

where $s_x(\sigma, \tau)$ is the sample standard deviation of the transformed values of texel aspect ratios back-projected onto a surface with orientation, (σ, τ). Again, the constants are collected into a single term, k, which is irrelevant to the estimation problem. The final set of derivatives in Eq. (39) are too cumbersome to write out here, but they are relatively easy to derive. Note that the prior term which appears in the likelihood function, $1/s_x(\sigma, \tau)^n$, punishs interpretations which correspond to broad distribution of back-projected, transformed aspect ratios.

4.3. Position

The natural default prior for surface texel positions is the one proposed by ref. [8] (in a slightly different theoretical context) and used by ref. [10] in their ideal observer work: that the fixed number of texels appearing in an image were 'dropped' onto a surface uniformly and independently. The underlying probability for position in the plane of the surface is therefore uniform over the area of the surface; that is, it is given by

$$p(u, v) = \tfrac{1}{A} \tag{40}$$

where A is the area of the surface. The likelihood function for texel position information derived from this process is simply given by

$$p(\{x_i, y_i\}|\sigma, \tau) = \frac{1}{A(\sigma, \tau)^n} \prod_{i}^{n} \frac{\partial u_i}{\partial x_i}\frac{\partial v_i}{\partial y_i} \tag{41}$$

where $A(\sigma, \tau)$ is the back-projected area of the image, exactly the ideal observer derived by Kanatani and by Blake et al. for texel position information. Note the appearance of the back-projected area in the prior term. This serves the same role for the generic texel position estimator as the average back-projected length served for the generic texel length estimator—it renders the estimator scale-invariant.

5. Simulations

I ran a number of simulations to analyze the information content of texture patterns and to test the efficiency of the generic ideal observers. For most of the simulations, I used texture patterns composed of random arrays of ellipses. These have the advantage that the moment formulation of texture information is complete for these textures—ellipses on surfaces project to ellipses in the image and the shapes of ellipses are completely characterized by their second-order moments. I also applied the models to analyze the information content of a particular, naturalistic class of synthetic textures; the Voronoi textures introduced by [27] into psychophysical work. Fig. 1 shows examples of both types of textures.

The simulations are organized into four sets. The first set is used to sketch out the basic structure of texture information for surface orientation. They determine lower limits on the reliability of texture patterns and analyze the reliability of Voronoi textures, as a special class of constrained texture patterns. The second set of simulations is designed to test the generalizeability of the generic observers by computing their efficiency for estimating surface orientation for a variety of different classes of surface textures. The third simulation set analyzes the role of field of view in determining texture cue informativeness, looking not only at how texture cue reliability changes with field of view size, but also at how texture cue informativeness depends on position in the image. The final set of simulations studies the role of the isotropy constraint in determining the reliability of foreshortening information.

5.1. Methods

In order to measure the reliability of the information provided by each of the three texture cues about planar surface orientation, I simulated three different ideal observers which, given a description of the texels contained in a texture pattern, estimated the orientation of the underlying surface from one or the other of the cues. The ideal observers were modeled as maximum likelihood (ML) estimators of surface orientation based on each of the three marginal likelihood functions described in Section 3. The ideal observers were derived with complete knowledge of the prior structure of the ensembles on which they were tested (using Eq. (5) and the projection model described in Section 3). The standard deviation of each ideal observer's surface orientation estimates, derived from Monte Carlo simulations, served as a measure of the reliability of the information provided by that cue. The measure is akin to the Fisher information measure, which specifies a theoretical lower bound on the standard deviation of an estimator[5] ([24]). Both provide exact bounds on an estimator's standard deviation in the limit as the number of independent texture measurements goes to infinity.

Fisher's measure of statistical efficiency provides the standard comparator for measuring the efficiency of the generic observers. Absolute efficiency according to this formulation is given by the ratio of variances of two estimators, one being the true ideal observer for a particular ensemble of texture stimuli. For slant, this would be

$$E = \frac{\mathrm{Var}_{\mathrm{ideal}}[\sigma]}{\mathrm{Var}_{\mathrm{gen}}[\sigma]} \tag{42}$$

where $\mathrm{Var}_{\mathrm{ideal}}[\sigma]$ is the variance of slant estimates obtained from the ideal observer and $\mathrm{Var}_{\mathrm{gen}}[\sigma]$ is the variance of slant estimates obtained from the generic observer. The true ideal observers for an ensemble are distinguished from the generic observers by the fact that they have complete knowledge of the ensemble statistics. The generic observers may have only partial or incorrect knowledge.

All of the simulations mimicked a condition in which observers viewed a textured surface through a rectangular window fronto-parallel to the viewing direction. The size of the window was $25° \times 25°$ of visual angle, except for the simulations in which I systematically varied the field of view of the observer. The images were constrained to contain 150 independent texels. Each stimulus used in the simulations was generated in three steps. First, a surface texture was randomly sampled from a pre-defined ensemble of textures (corresponding to a particular setting of model parameters). Second, the texture was 'rendered' for a particular surface slant and tilt using the spherical projection variant of the projection model described in Section 3. Finally, the rendered stimulus was cropped by the spherical image of the rectangular window described above.

The textures in Fig. 1 are examples of the stimuli as they would appear rendered on a computer screen, with the addition of a simulated window frame and a lower lip on the surfaces, for realism.

5.2. Simulation set 1: basic properties of texture information

5.2.1. Example likelihood functions

Fig. 3 shows example likelihood functions for each of the three cues, computed for the image shown in figure la. This image is drawn from a highly constrained ensemble of surface textures, much like those used in the psychophysical experiments described in the companion article [5]. The computed likelihood func-

[5] Fisher information is an approximate upper bound on the value of $1/\mathrm{Var}[\hat{S}]$ for any tmbiased estimator of S, thus it also provides a lower bound on the standard deviation of an estimator.

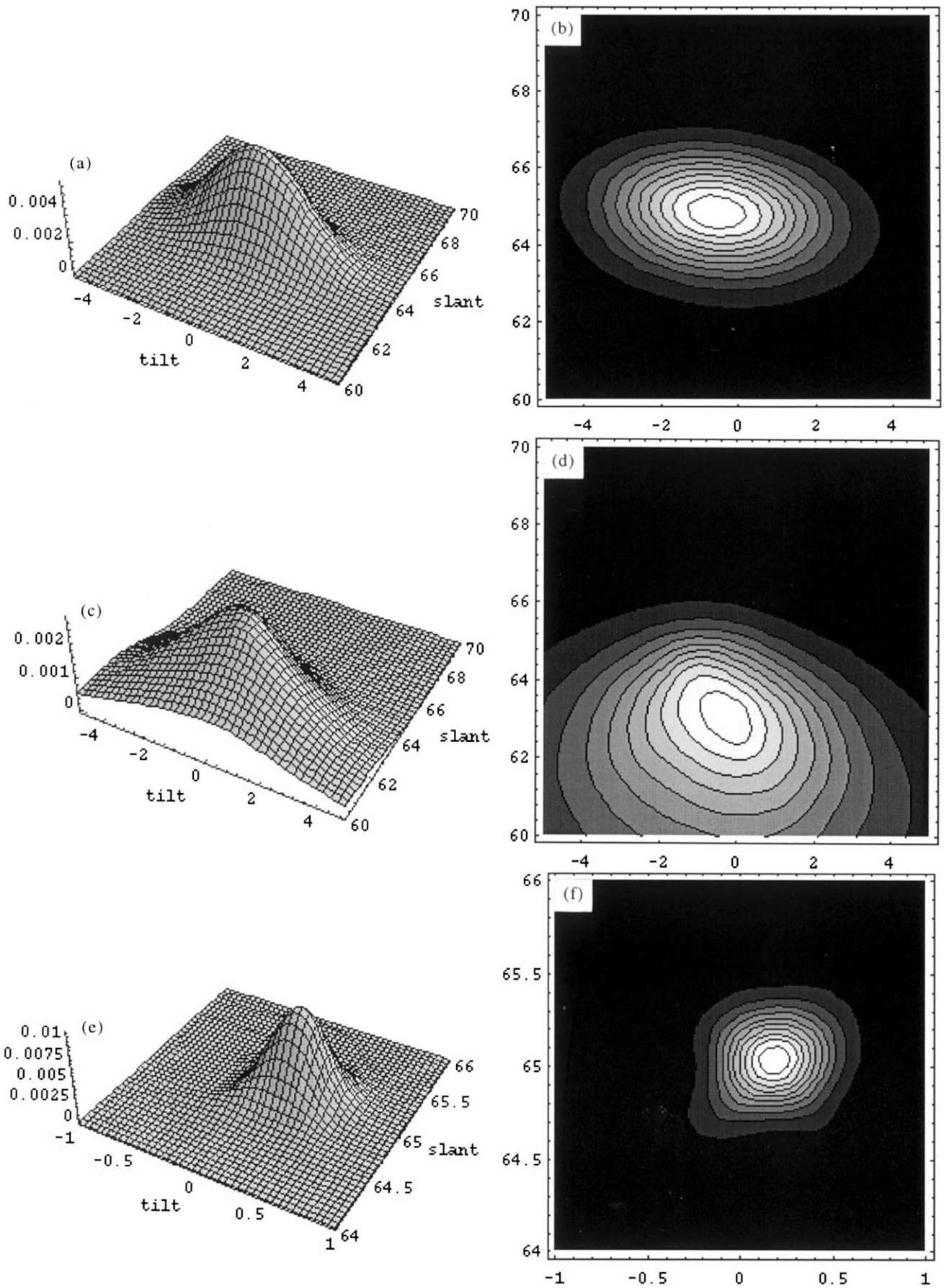

Fig. 3. Plots of likelihood functions for a specific texture stimulus, in this case, rendered with a slant of 65° relative to the fronto-parallel plane. (a) and (b) show plots of the likelihood function for the scaling cue, (c) and (d) for the position cue and (e) and (f) for the foreshortening cue.

tions were derived with complete knowledge of the statistics of the ensemble of surface textures from which the sample was drawn. The plots provide an illustration of the relative informativeness of the different texture cues about surface orientation, for the defined ensemble of textures. The spreads of the three distributions, assuming the stimuli are representative of the texture ensemble, give an idea of the reliability of the information about surface orientation provided by each cue.

As reflected in the orientation of the likelihood functions, none of the three texture cues independently specifies surface slant and tilt. Thus, knowledge of one generally helps to estimate the other. Simulations, however, show the correlation between the two surface orientation parameters to be very weak. Moreover, both negative and positive correlations between slant and tilt are found for different samples drawn from the same ensemble and there is no systematic correlation between estimates of slant and tilt, meaning that independent information about slant or tilt will not bias estimates of the other.

5.2.2. Simulation 1a: lower bounds on texture cue informativeness

Consider the following scenario. You are told that you will view a series of images of textured surfaces oriented away from you at some fixed slant and tilt. These images have a fixed size and always contain a fixed number of texels. You are further informed that the surface textures from which the textures were projected conform to the qualitative constraints of homogeneity, isotropy and independence of texture elements. Finally, you are asked to set a lower limit on the reliability of the texture cues contained in the images. In order to solve this problem, you would need to search the space of all surface texture ensembles which match the specified constraints and find the one(s) for which the ideal observers have the worst performance. One might suspect that the worst possible performance would be equivalent to guessing, but as we shall show, this is not the case.

Searching the space of allowed texture ensembles is a daunting task, however, one can take advantage of the observation that the surface texture ensembles which lead to the least informative images will, in some sense, be the most irregular. I therefore conjecture that the surface texture ensembles leading to the least informative stimuli will be ones with maximal entropy, in some sense. In order to set bounds on the reliability of the three texture cues, therefore, I simulated ideal observers for surface textures generated from the same maximal entropy priors used to derive the generic observers, with the exception of the prior model for texel shape. For texel shape, one can take advantage of the fact that it is a two-dimensional entity and reduce its informativeness by removing one of the degrees of freedom in texel shape. Since orientation is intrinsic to all possible texture elements, we can only remove the shape dimension corresponding to aspect ratio, by using line element textures.

For the position and foreshortening cues, the maximal entropy priors we end up with are exactly those used by Blake et al. for their analysis of the information content of textures; thus, their information measures provide lower limits on cue reliability; that is, they set the worst-case limits on cue reliability. To my knowledge no one has performed a similar analysis for scaling information. As pointed out in the generic observer section, however, one can define a maximal entropy prior distribution for texel length—the exponential distribution, with a randomly selected scale constant. This prior defines the surface texture ensemble giving the least reliable texel size information[6].

Fig. 4 shows the standard deviations of ideal observer estimates of surface slant and tilt for the maximum entropy distributions of surface texel parameters. For the particular field of view used ($25° \times 25°$), scaling information is the least reliable, while the relative reliability of the position and foreshortening cues depends on surface slant. As pointed out by Blake et al., however, the ordering of the cues varies as a function of field of view. They showed that the relative reliability of foreshortening and position cues varied with the size of the field of view: foreshortening dominated for small fields of view, but position dominated for large fields of view. We will take up this point further in the section on field of view effects.

Note that in this worst case scenario, none of the texture cues provide reliable constraints on surface slant and tilt for surfaces slanted away at a small angle, but they become much more useful cues at large slants.

5.2.3. Simulation 1b: voronoi textures

Rosenholtz and Malik (1994) [27] introduced into shape from texture studies a synthetic model of surface textures which generates highly naturalistic tiled textures (see Fig. 1b). This section analyses the statistical structure of such textures and shows that one can derive approximate ideal observers for them. We do this partly to show that the ideal observer formulation can be broadly applied beyond textures composed of simple elliptical elements and partly as a case study into what the relative reliability of texture cues would be for textures in which different texel properties (shape and size) emerge naturally from a single underlying texture generation process. For Voronoi textures, the ideal observers are approximate in the sense that they use only second-order spatial moments of the polygonal 'tiles' which make up the textures. Other, higher-order spatial moments could, in theory, provide further information.

[6] One might suggest using a maximal entropy prior defined for some transformation of length, (e.g. log(length)). We have, in fact, tried a variety of 'broad' priors on texel length, including uniform priors, and found that none of them led to worse ideal observer performance than the exponential model.

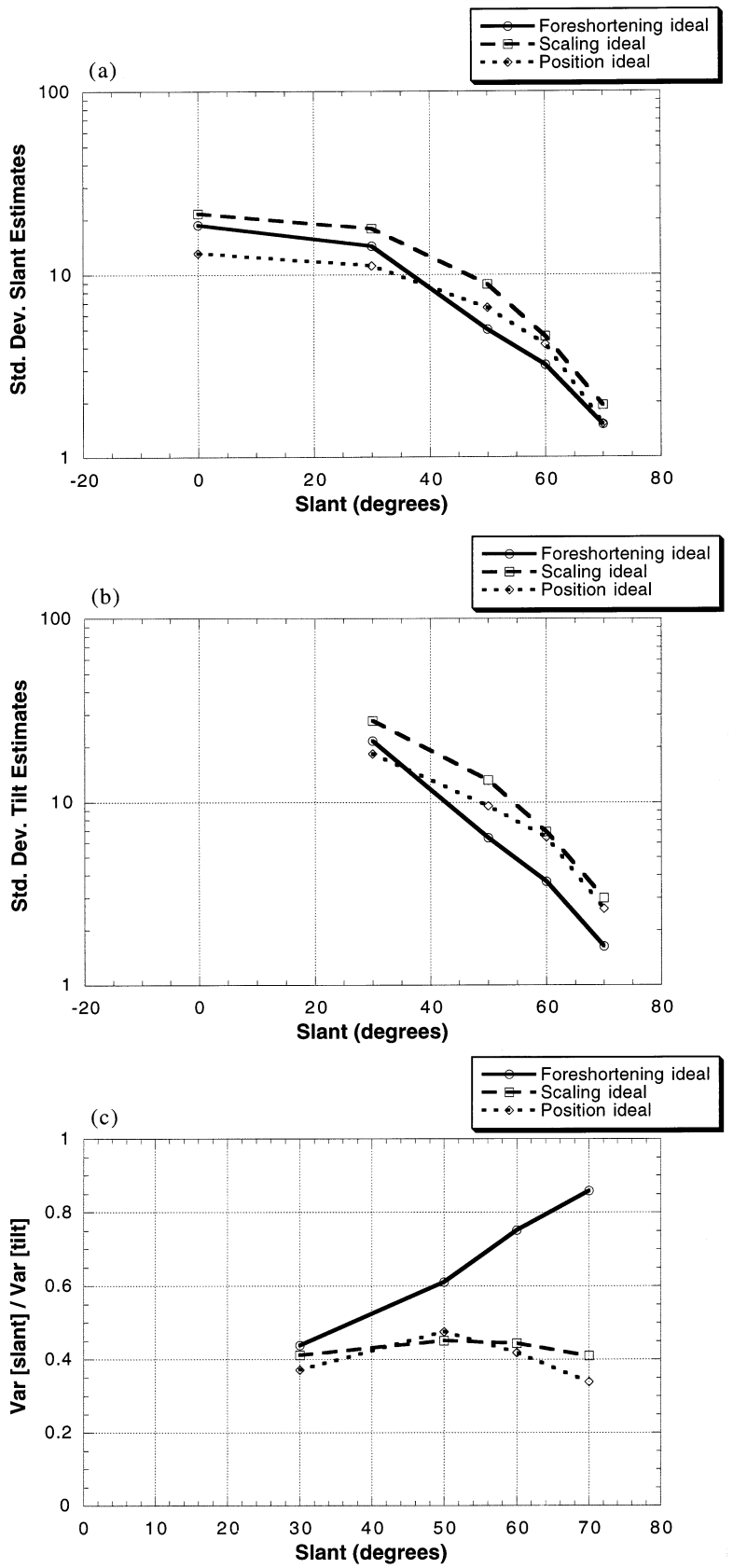

Fig. 4. Plots of the standard deviations of ideal observer estimates of surface slant (a) and tilt (b) for the stimuli used in the the first set of simulations. These measures represent worst-case scenarios for the viewing conditions used in the simulations (25° × 25° field of view, 150 texels). (c) shows the relative reliability of the information provided by the three texture cues about slant and tilt, expressed as the ratio of variances of the two estimates.

The textures introduced by Rosenholtz and Malik are comprised of the polygons formed from a Voronoi diagram [25], created by drawing polygons around texel positions in the plane in such a way that each polygon contains all points in the plane which are closer to a particular texel position than to any other. I analyzed the statistics of two classes of Voronoi textures; one generated from an unconstrained model of texel positions (a 2-D Poisson process), the other generated from a more constrained reaction-diffusion model [26], similar in spirit to a texture generation model used by [27]. The model simulated a continuous, repulsive inhibitory field around each texel position which governed the migration of texels over time. Samples of the process drawn when it was in a steady state provided constrained random lattices for generating surface textures.

Within the moment-based formulation described here, I represent each polygon in a Voronoi texture by its second-order moments of inertia. Treating the resulting moment tensor as the matrix for the quadratic form of an ellipse, one can compute its eigen-vectors and eigen-values to derive the length, orientation and aspect ratio of the polygon. Doing this for every polygon in a large set of textures provides sample data to fit prior distributions for the texel attributes used by the ideal observers. Inserting these prior distributions into the general ideal observer formulation gives appropriate ideal observers for the scaling and foreshortening cues contained within Voronoi textures. We measured several sets of correlations to test the independence assumptions built into the ideal observer formulation. These included correlations between the sizes, orientations and aspect ratios of individual texels, as well as correlations between the same properties of neighboring texels. The only non-zero correlation found was between the lengths of neighboring texels. Even this correlation, however, was quite low ($\rho = 0.11$ for the constrained Voronoi textures and $\rho = 0.08$ for the unconstrained textures) and dropped off quickly to zero when more distant neighbors were considered. The results support approximating the texel properties as being independent.

The derived ideal observers were run on texture descriptions computed from real Voronoi textures. I further tested the fit of the derived ideal observers by running them on elliptical element textures generated with the same first-order statistics as calculated for the Voronoi textures. These simulations gave essentially identical results to the simulations using real Voronoi textures, further validating the model. Fig. 5 shows the results of the simulations. The pattern of relative reliability between the cues is the same for both texture ensembles, with the foreshortening cue being most reliable at all but the lowest slants.

5.2.4. Discussion

The most striking feature of the simulation results is the marked increase in reliability of texture information with increasing surface slant. The behavior holds regardless of the surface texture ensemble used to generate stimuli and reflects the fact that the reliability of texture information depends inherently on surface slant. This result replicates a similar finding reported by Blake et al. in their ideal observer work [10]. A second property evident in the graphs which seems to hold generically for all surface texture ensembles is that texture provides a more reliable cue to surface slant than it does to surface tilt. Malik and Rosenholtz [23] also found this to be the case when running their estimator on natural textures [27]. Note also that the patterns of relative variance in estimates of slant and tilt are qualitatively similar across the texture ensembles tested here and was also borne out in numerous simulations with other ensembles. This is not surprising since the relative reliability of the information provided by texture cues about slant and tilt is determined by the geometry of the viewing situation, not the underlying statistics of the texture ensembles.

It is also of interest to note the ordering of cue reliability for the Voronoi textures. For both ensembles simulated, foreshortening information was best, followed by scaling and then by position. In Voronoi textures, the variability in the sizes, shapes and positions of surface texels, which determines the reliability of the three texture cues, is highly correlated, since all three properties derive from a common generator function (the position generator). It seems likely that the same will be true of most natural textures; thus, we expect that the result has some degree of generality and is probably more reflective of the behavior of natural textures than the results obtained with the maximal entropy texture models.

5.3. Simulation set 2: ideal and generic observers compared

In section 4, I derived a set of generic observers for each of the three texture cues based on minimal assumptions about the ensembles of surface textures for which they might be used. In this section, I measure the efficiencies of the three generic observers when tested on stimuli generated from a wide range of surface texture ensembles, with varying degrees of regularity. In many cases, the prior statistics of the ensembles deviate significantly from the unconstrained priors assumed in the derivation of the generic observers. The true ideal observers differ from the generic observers in that they are derived with complete knowledge of the test ensemble statistics; thus, they provide absolute limits on performance of any estimator.

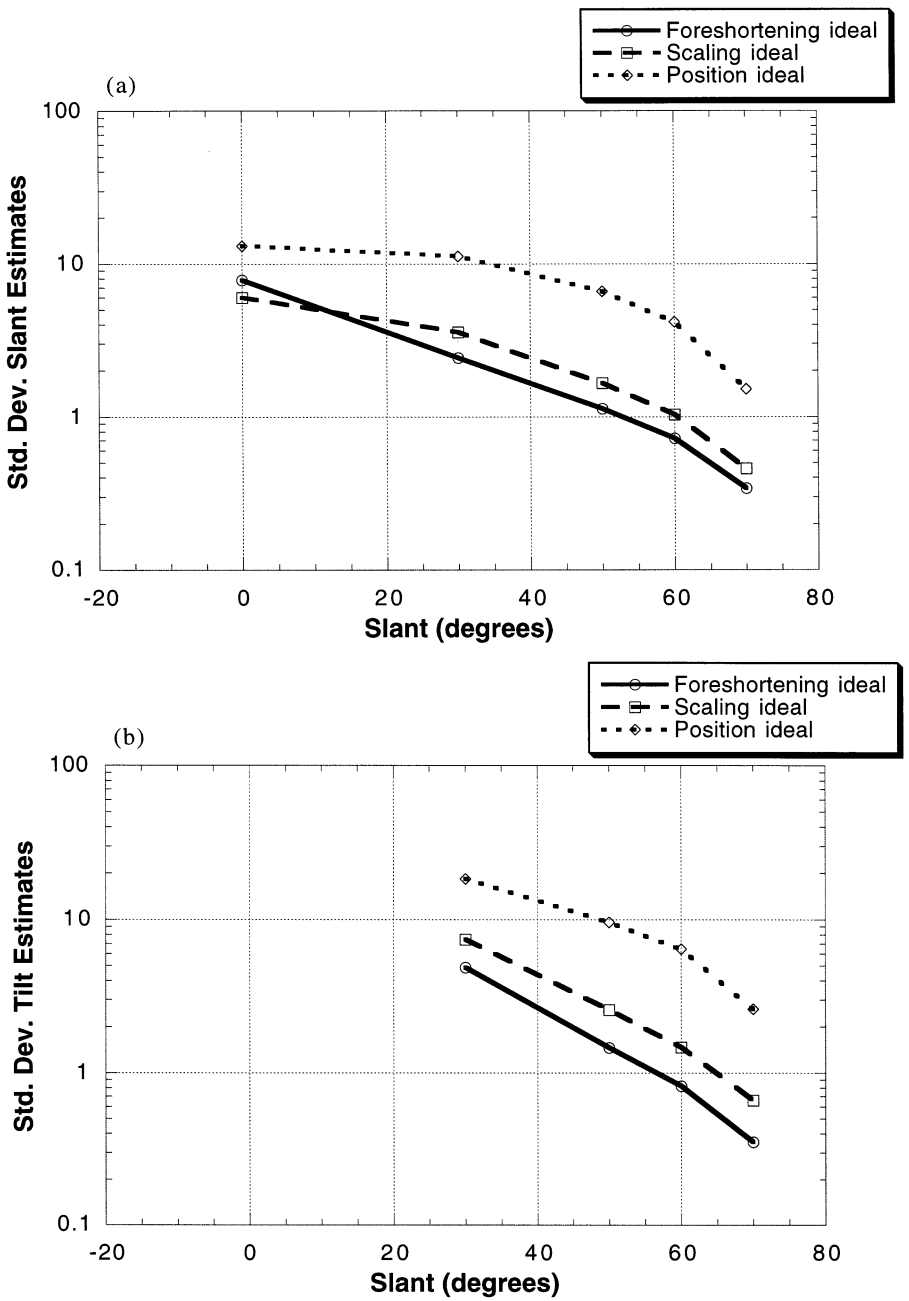

Fig. 5. Reliability data for the three texture cues for two types of Voronoi textures. (a–c) shows standard deviations of the slant and tilt estimates and the relative reliability of slant and tilt estimates as derived from each of the three texture cues for the unconstrained Voronoi textures (generated from a 2D Poisson process). (d–f) shows similar data for the constrained Voronoi textures (generated from the constrained reaction-diffusion process).

The section contains three sets of simulations. In the first, I measured the efficiencies of the three generic observers for the constrained ensemble of Voronoi textures used in simulation 1b. The other two sets of simulations analyze the performance of the foreshortening and scaling generic observers on broader ranges of texture stimuli. For conciseness, I will only describe the results for slant estimation. Those for tilt were qualitatively the same.

5.3.1. Simulation 2a: voronoi textures

I ran Monte Carlo simulations of the generic observers for the Voronoi textures used in simulation 1b. Fig. 5a shows the standard deviations of slant estimates derived for an ideal observer which had exact knowledge of the prior distribution of texel shape, size and position in the Voronoi textures (estimated from samples of the texture process). These distributions differed from and were significantly more constrained than

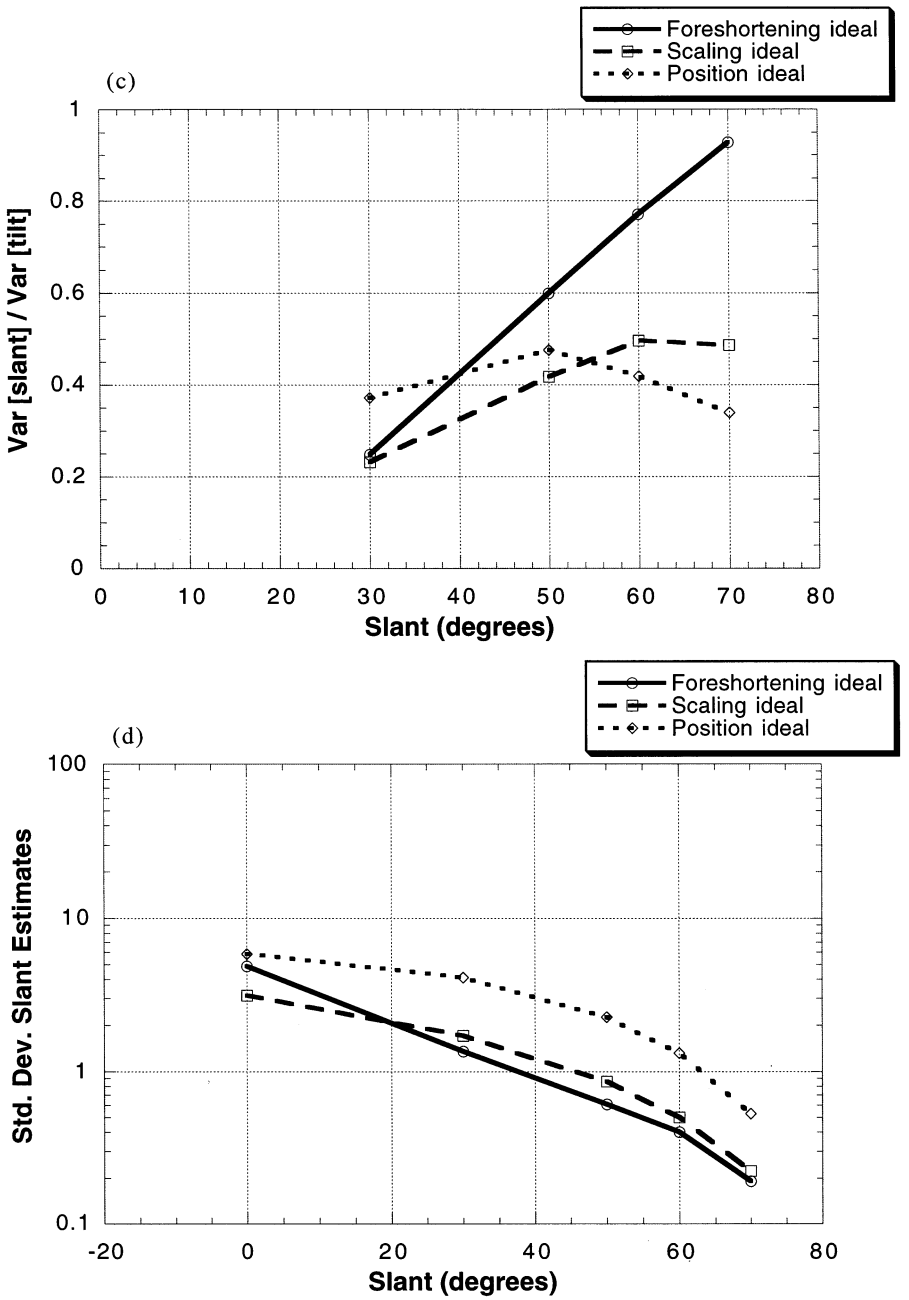

Fig. 5. (*Continued*)

those used to derive the generic observers[7]. Despite these differences, the standard deviations of the generic scaling and foreshortening observers' estimates of slant deviated only a little from those of the true ideal observers. This is reflected in the high efficiencies for these observers shown in Fig. 6.

The efficiency of the position generic observer was much lower than that of the other observers. I suspect that this results primarily from the fact that the generic position observer (which is really a 'density' estimator) does not take advantage of the isotropy constraint. Images of textures constructed from constrained lattices of texel positions carry information in the relative positions of neighboring texels, with the local compression of distances in the tilt direction providing significant information when the lattice is constrained to be

[7] The distribution of texel lengths in the Voronoi textures was well fit by a Gaussian distribution. That of texel aspect ratios was well fit by a truncated (at 0 and 1) Gaussian. The distribution of texel positions was characterized by the energy function for the reaction–diffusion process used to generate random texel lattices, a much more constrained process than the 2D Poisson process assumed for the generic position observer.

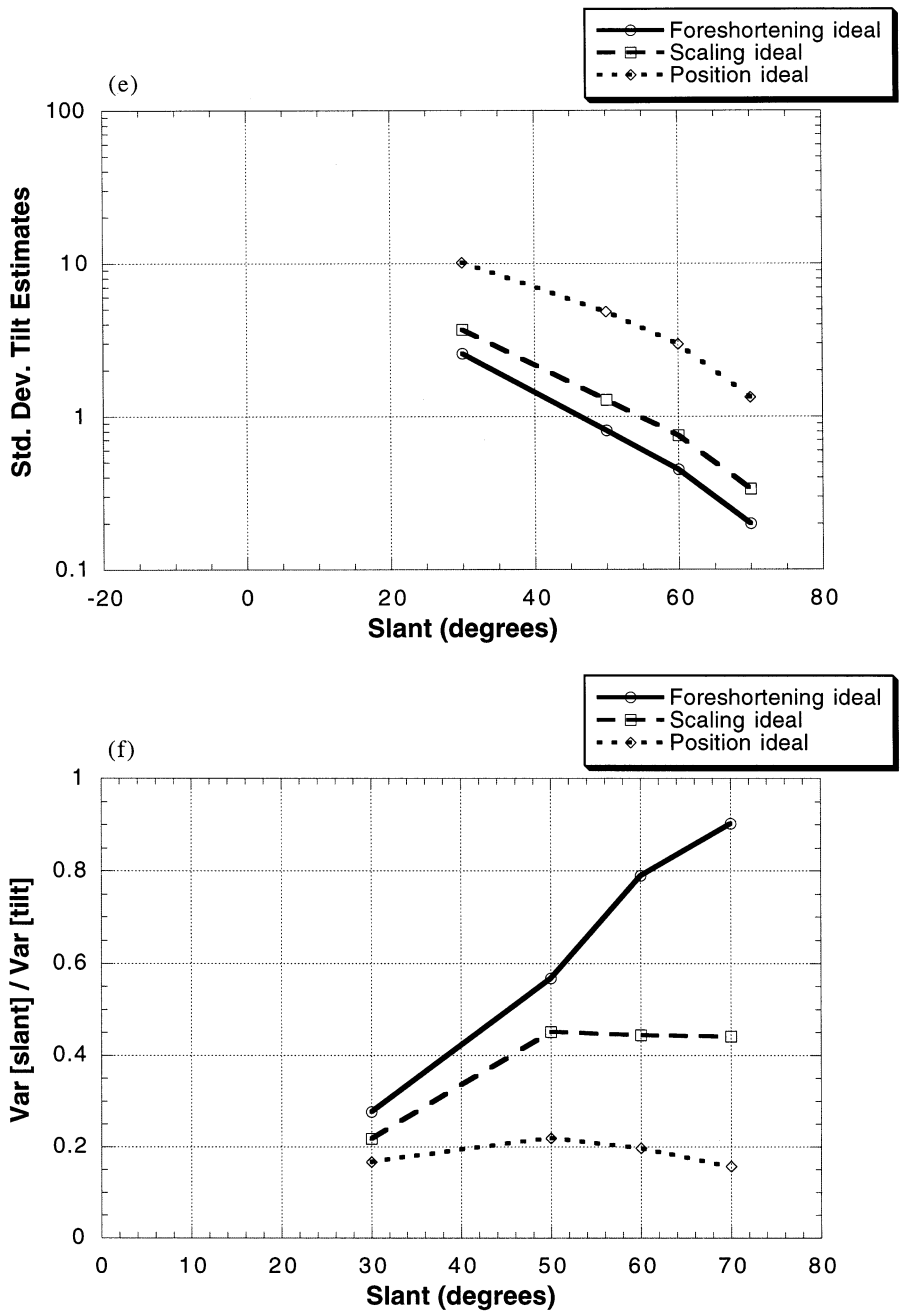

Fig. 5. (*Continued*)

isotropic. As we will see in section 5.5, the inefficiency of the generic position observer for images of constrained isotropic texture lattices is of approximately the same magnitude as the inefficiency of a foreshortening ideal observer which does not incorporate an isotropy constraint.

5.3.2. Simulation 2b: scaling

I tested the generic observer for scaling information on textures whose texel lengths were drawn from a Gaussian distribution. I measured generic observer efficiency for a range of prior distributions of texel lengths, ranging from low to high variance distributions. The absolute ideal observers for these ensembles were derived assuming no prior knowledge of the absolute mean of the length distribution. Fig. 7 shows plots of generic and ideal observer performance as a function of the standard deviation of surface texel sizes. As expected, the performance of both observers gets worse as the variance is increased. More notably, the generic observer performed almost as well as the ideal observer over the entire range of texture ensembles tested, with the lowest efficiency measured being 81%. The result is particularly striking given the large discrepancy be-

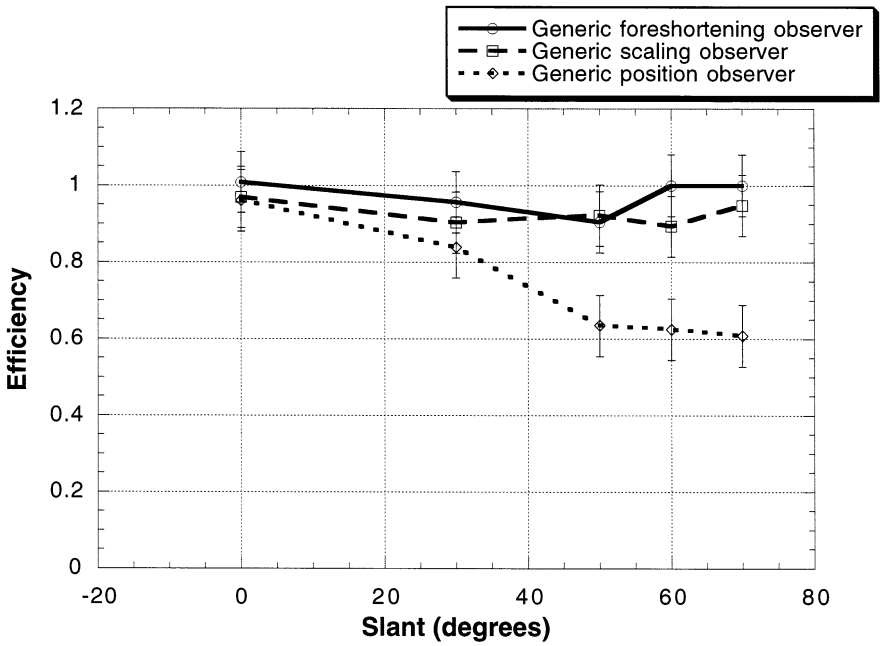

Fig. 6. Absolute efficiencies of the generic observers as a function of surface slant for the constrained class of Voronoi textures used to generate stimuli for simulation 1b.

tween the Gaussian distributions of texel lengths in the test ensembles and the exponential distribution used to derive the generic observer. This hints at the general applicability of the generic scaling observer derived here.

5.3.3. Simulation 2c: foreshortening

In an initial test of the generic foreshortening observer, I measured the efficiency of the observer for ensembles of textures whose texel aspect ratios were drawn from specific distributions in the family of distributions used to derive the generic observer. The difference between the generic observer and the absolute ideal observer for these ensembles was that the absolute ideal had prior knowledge of the parameters defining the prior distribution for each test ensemble, while the generic observer did not. The efficiency of the generic observer was above 80% for all of the test ensembles used. The appropriate way to read this result is that, within the context of the family of aspect ratio distributions used to derive the generic observer, prior knowledge of the specific distribution used to generate texel aspect ratios is not critical to the information specified by foreshortening.

A better test of the generalizeability of the generic foreshortening observer is to measure its efficiency for ensembles of textures whose texel aspect ratios are drawn from distributions not included within the defined family of distributions. Fig. 8 shows the efficiency of the generic observer for test ensembles whose texel aspect ratios are drawn from Gaussian distributions with a range of aspect ratios (with a mean aspect ratio of 0.5). The generic observer's efficiency was generally high (above 80%) in simulations using a wide range of other Gaussian distributions of aspect ratios, though it dipped as low as 64% in a few cases.

The results of the simulation set suggest a certain degree of generalizability of the generic foreshortening observer derived here. One can, however, easily create 'strange' prior distributions of aspect ratios which will lead to lower efficiencies. For example, a bimodal distribution of aspect ratios with narrow peaks around 0.3 and 0.7 led to an efficiency of only 9%. High efficiency in this case would clearly require prior knowledge of the peculiarities of the specific aspect ratio distribution for the texture ensemble, something which is unlikely to be available in a general setting.

5.4. Simulation set 3: field of view effects

Increasing the field of view (FOV) on a texture pattern increases the information available for estimating planar surface slant. Two factors underlay this effect. The first is a trivial one: increasing the field of view increases the amount of texture visible. The second and more interesting factor is the changing geometry of local texture distortions as a function of position in the image. One can isolate this factor by considering image textures with a fixed number of texture elements, independent of field of view size.

The main dimension of interest in the image is the direction of surface tilt (vertical, for images of ground planes). Changes in field of view size in the perpendicular direction have only small effects on the information

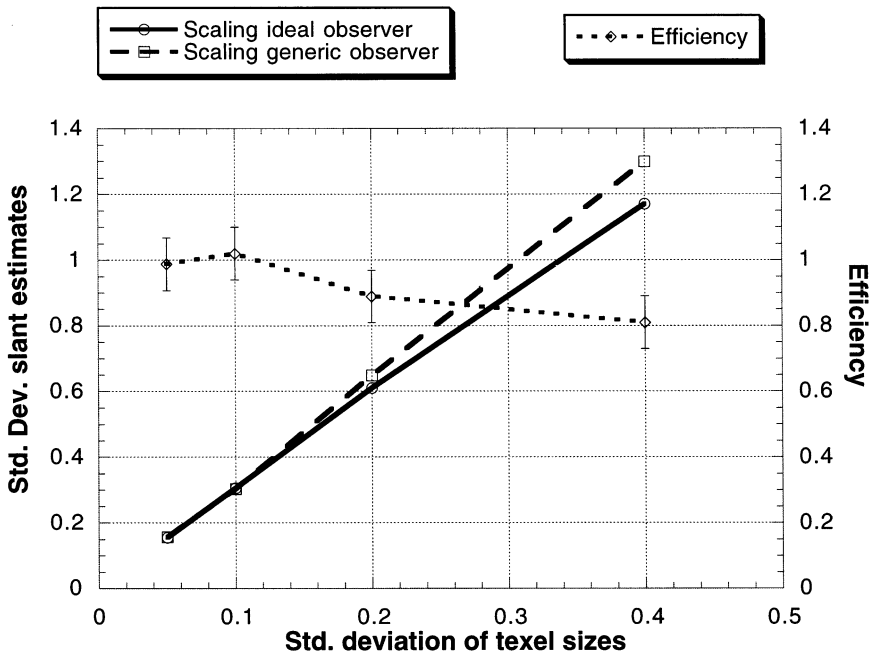

Fig. 7. Standard deviations of slant estimates for ideal and generic scaling observers as a function of the standard deviation of texel lengths on a surface (expressed as a proportion of the mean texel length). The slant of surfaces used in these simulations was fixed at 65°. Texel lengths were Gaussian-distributed, truncated at zero. The absolute ideal observer had full knowledge of the true prior for each texture ensemble, with the exception of the mean length. The dotted curve shows the absolute efficiency of the generic scaling observer, which can be read off of the right-hand axis.

content of texture patterns. For simplicity, I will consider field of view effects on the problem of slant estimation for surfaces with a vertical tilt. The results and discussion do, however, generalize to the estimation of tilt as well as slant.

Blake et al. (1993) [10] have shown that increasing the field of view on planar, slanted surfaces increases the reliability of both texture density and texture foreshortening information, even when the number of texels (or texture measurements) remains fixed. This makes intuitive sense—we tend to think of texture information being carried by differences in texture properties in 'near' and 'far' parts of an image. It therefore stands to reason that increasing the field of view, which increases the total range of these differences, should increase texture reliability. This simple observation, however, is complicated by the fact that the texture information from different parts of an image is more or less effective in determining surface orientation estimates. The first few simulations described in the paper make clear the fact that texture cue reliability increases with increasing surface slant. Thus, a window on the upper half of a ground plane will provide more reliable information about surface orientation than a similar sized window on the bottom half, even if the number of independent texture measurements available from the two windows is equal. This is because the local slant of points on a ground plane increases with height in the image (to 90° at the horizon).

To illustrate the point I performed a set of three simulations using the textures with the same statistics as the constrained Voronoi textures introduced in Section 5.2.3. In the simulations I began with a 5° FOV image of a surface slanted away from the line of sight at an angle of 65° vertically (measured relative to the line of sight from the fovea). I then increased the field of view in three different ways; symmetrically around the center line, strictly downwards and strictly upwards. The results are shown in Fig. 9. The greatest advantage clearly derives from the upward expansion of the field of view. The downward expansion does not improve the scaling cue and leads to an actual decrease in performance for the foreshortening cue.

The results shown in Fig. 9 can be explained by noting how texture information is redistributed in the simulations as the field of view is expanded in different directions. As the field of view is expanded upward, the texture information is distributed over regions of the surface with larger local slants, while the opposite is true as it is expanded downwards. For the downward expansion, the distribution of texture into regions of the image which are less informative mitigates the advantages gained for the scaling and position cues by having a larger field of view. For the foreshortening cue, which is inherently local, the redistribution of texture information makes the cue less reliable, as its reliability depends simply on local surface slant.

In these simulations was held the number of texels constant across changes in field of view size. In natural

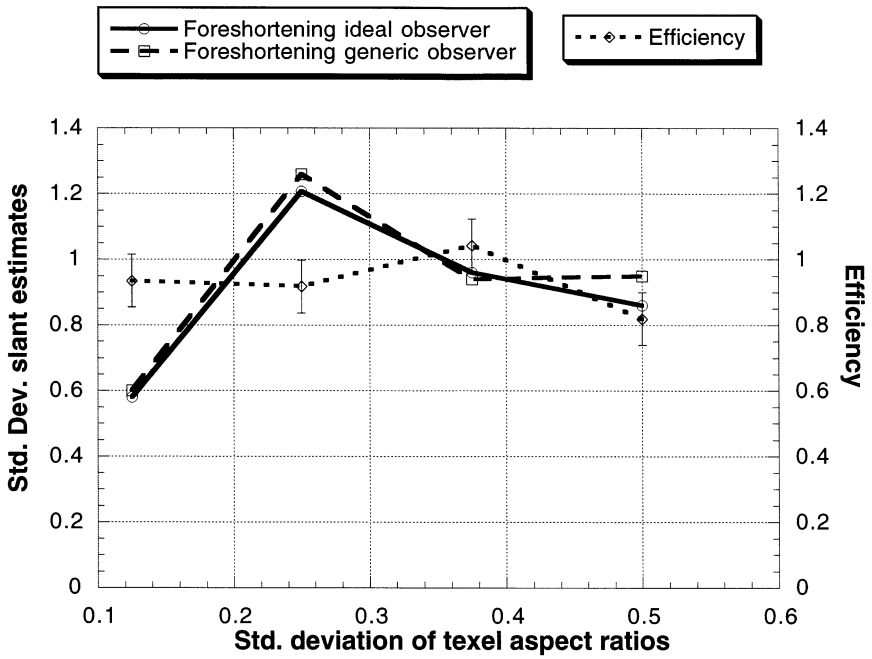

Fig. 8. Standard deviations of slant estimates for ideal and generic foreshortening observers as a function of the standard deviation of texel aspect on a surface. The slant of surfaces used in these simulations was fixed at 65°. Texel aspect ratios were Gaussian-distributed, truncated at zero and one. The absolute ideal observer had full knowledge of the true prior for each texture ensemble. The dotted curve shows the absolute efficiency of the generic foreshortening observer, which can be read off of the right-hand axis.

viewing, where increasing the field of view leads to more texture data, it should improve estimation performance. The results described here, however, point to the fact that the optimal window on a planar surface is over the furthest part of the surface. Looking at the horizon is, in the limit, the most informative view one can have about surface orientation for surfaces known to be planar. In the real world, an observer must trade off this consideration with the decreased likelihood that the planarity constraint will hold over large distances.

5.5. Simulation set 4: the importance of the isotropy constraint

Up to this point in the analysis, I have only considered isotropic surface textures. Many models of shape-from-texture assume isotropy, but it is not necessary to do so to use the foreshortening cue for estimating surface shape or orientation under perspective projection. The amount by which a texture pattern is locally foreshortened changes with position in the image, an effect which derives from the fact that at any given point in the image, the orientation of a surface relative to the local line of sight depends both on the global orientation of the surface and the angle of the line of sight to that point. The resulting 'gradients' in foreshortening provide information about surface slant even without prior knowledge of isotropy. Malik and Rosenholtz [11], for example, have developed a computational model which uses spatial differences in the texture distortion induced by perspective to estimate surface shape. Such a model would be particularly useful when applied to textures which are not, in fact, isotropic.

This raises the question of how sensitive the foreshortening cue is to prior knowledge of surface texture isotropy. To answer that question I compared the performance of a sub-ideal observer which does not assume isotropy to an ideal observer which has prior knowledge of isotropy. In particular, the 'foreshortening-without-isotropy' ideal observer assumes that surface textures could be uniformly stretched or compressed by some global affine transformation before projection into the image. The affine transformation is parameterized by two values: the direction and amount of compression. These two parameters characterize the global 'shape' of a texture pattern. The foreshortening-without-isotropy ideal observer simultaneously estimates both the orientation of a surface and the global shape parameters of its surface texture.

Fig. 10 illustrates the uncertainty induced by discarding knowledge of isotropy. The figure shows a plot of the foreshortening cue likelihood function as a function of surface slant and the amount of surface texture compression. The tilt here is fixed at vertical and the direction of compression is in the direction of tilt. The likelihood function was derived from a stimulus pattern similar to those used in simulation 1b. The plot shows a strong correlation between estimates of surface texture compression and surface orientation, even for the large field of view used to create the stimulus image. The strength of

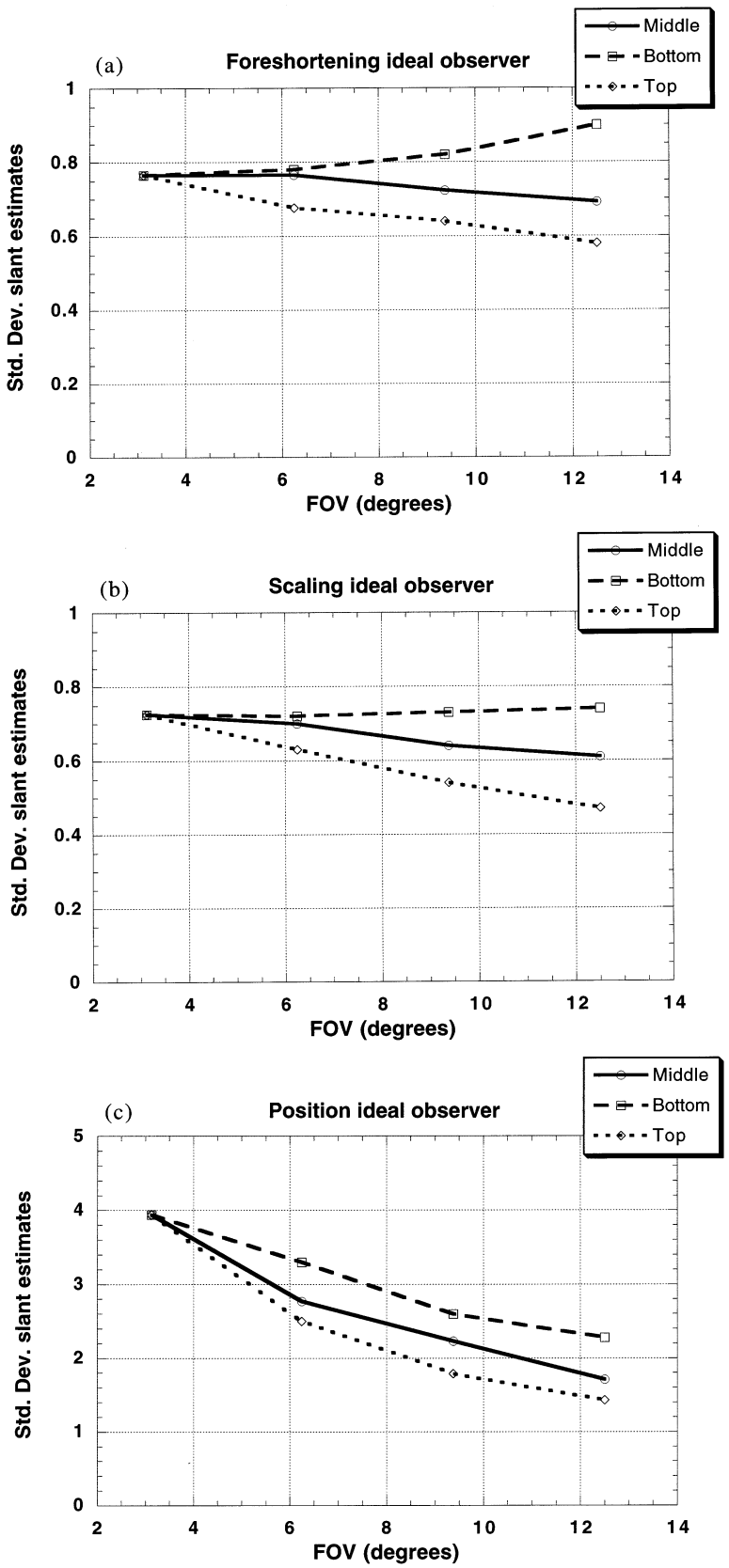

Fig. 9. Standard deviations of slant estimates for the three ideal observers as a function of field of view size. The field of view was varied in three different ways—by growing it symmetrically around the center of fixation, by growing it upwards, away from the center of fixation (towards the horizon) and by growing it downwards from the center of fixation. The simulated surfaces were at 65° from the fronto-parallel and contained 150 texels each.

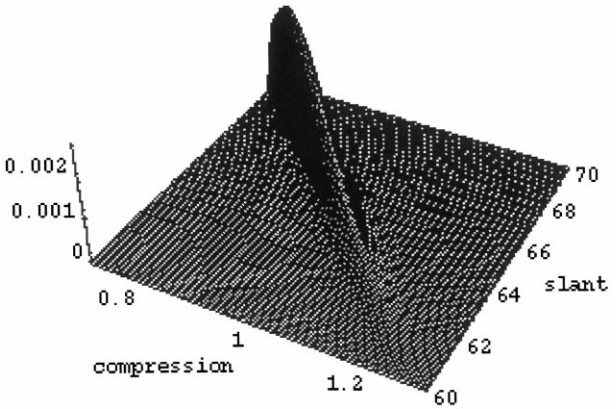

Fig. 10. The likelihood function generated by the image of a surface texture at 65° slant, expressed as a function of surface slant and amount of global texture compression. Smaller values of compression correspond to the surface texture being stretched in the direction of tilt; thus, for a given image texture, an estimator which selected a small value of texture compression would have to compensate by estimating the slant to be larger than it is. The stimulus was generated using the same texture parameters and viewing parameters as used for the simulations described in the text.

the correlation reflects the importance of the isotropy assumption in constraining estimates of surface orientation.

In order to quantify the power of the isotropy constraint, I ran the foreshortening-without-isotropy ideal observer and the true ideal observer on stimuli generated from isotropic surface textures (using the same set of stimuli used in Section 5.2.3). Fig. 11 shows the results of the simulation. Removing prior knowledge of isotropy increased the standard deviation of observer estimates by roughly 4-fold for slants much greater than zero. We have replicated this effect with a number of other texture ensembles. Decreasing the field of view size should lead to an even greater decrement in performance between the true foreshortening ideal observer and the foreshortening-without-isotropy ideal observer—for isotropic surface textures. Were surface textures to be generated by a process which induces random amounts of anisotropy (texture compression), of course, the foreshortening-without-isotropy ideal observer would be the true ideal observer and would perform better than the foreshortening ideal observer which incorrectly assumed isotropy.

6. Conclusions

I have derived a general formulation for ideal observers for estimating planar surface orientation from texture for three different texture cues—perspective scaling, foreshortening and position. The formulation extends previous ideal observer work in the area in two important ways. First, it includes the ideal observer for perspective scaling information. Second, it generalizes previous formulations of the foreshortening ideal observer from simple line element textures to textures composed of spatially extended elements. The current

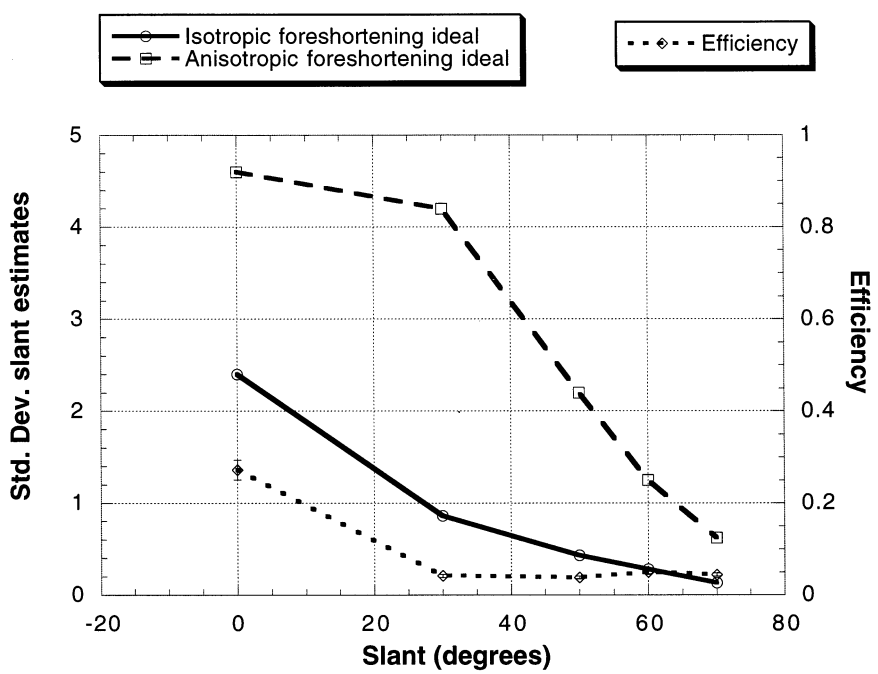

Fig. 11. Plots of the standard deviation of ideal observer estimates of surface slant for stimuli like those used in simulation 1, as a function of surface slant. The two ideal observers shown are ones which use foreshortening information, either with or without prior knowledge that the surface texture ensemble was isotropic. The ideal observer which did not have such prior knowledge simultaneously estimated surface orientation and the amount and orientation of global surface texture compression. The dotted line shows the absolute efficiency of the foreshortening-without-isotropy ideal observer. The increase in efficiency at 0 degrees slant is probably due to the compressive nature of the slant scale.

formulation is more broadly applicable to region-based texture measurements such as spectral moments, while the previous formulation naturally lent itself to edge based measurements. The former is more robust, since it does not suffer from the sampling problems induced by arc-length foreshortening of contours.

I have further derived a set of generic observers for the three texture cues. Simulations have shown that the generic observers for perspective scaling and foreshortening are near-optimal estimators of planar surface orientation for a large range of texture ensembles. This result has two important implications. First, it shows that the generic observers may be broadly applicable in computer vision applications, when applied to real texture measurements. Second, it validates the results of the ideal observer analyses presented here, suggesting that they will generalize to the performance of other orientation-from-texture estimators.

In this concluding section, I will discuss the implications of the ideal observer analysis for human perception of planar surface orientation from texture. I will also discuss the implications of the results for estimating surface curvature of non-planar surfaces.

6.1. Implications of results

6.1.1. The effect of surface slant

Simulations show that all three texture cues become more reliable with increasing surface slant. The geometry of the slant-tilt representation suggests this result for tilt, since at low slants, large changes in tilt correspond to only small changes in the surface normal vector (small displacements on the Gaussian sphere). In the limit, as surface slant goes to 0°, tilt becomes undefined and hence completely ambiguous. The situation is more complicated for slant. The changing reliability of slant estimates derives from the geometry of the texture projection (see Section 5.4 for discussion).

In psychophysical experiments, we have found that human subjects' ability to discriminate surface slant from texture improves markedly with increasing surface slant, from 30 to 40° at 0° slant to 1–2° at 70° slant. This mimics the changing reliability of the information available for estimating slant. Together, the theoretical and psychophysical results suggest that the visual system should weight texture information relative to other cues differently as a function of slant. The size of the effect suggests that texture cues should be weighted more as surface slant increases, though this result depends on the results of a similar analysis of the information provided by the cues with which texture is combined, (e.g. stereo, motion, etc.).

6.1.2. Cue weighting

One of the issues that one would like to resolve from the ideal observer analysis is the relative informativeness of the different texture cues and how this varies in different viewing conditions. No simple resolution to this is available, since the reliability of the cues can vary independently of one another with changes in surface texture statistics. The question, then, is which are most reliable in images of natural textures—a question which is clearly beyond the scope of this paper. The only hint that we have regarding the relative reliability of the three cues is from the simulations of the Voronoi textures. In these textures, the statistics of the different surface texture properties on which the cues rely derive from a common generator function (the position generator). Thus these statistics co-varied, leading to a covariation in texture cue reliability across different ensembles of Voronoi textures. For the Voronoi textures, the relative reliability of the three cues remained fairly invariant over the different ensembles tested. One might well expect this to be the case for natural textures -rarely does one expect to find textures which have little variance in one or another shape property with high variance in another. For the Voronoi textures, the position cue was significantly less reliable than either the foreshortening or scaling cues, with the reliability of the latter two being similar, though that depended on the field of view (see below).

That the position cue is least reliable is consistent with the psychophysical literature [15,28,29] which suggests that position information (usually referred to as density information) receives little to no weight in perceptual judgments. Other psychophysical work [2,30,31] suggests that foreshortening information is weighted at least as heavily as scaling information and in most contexts tested, is weighted more heavily. This is consistent with the finding in our simulations that foreshortening information was somewhat more reliable than scaling for the textures simulated; however, the generality of this result must be qualified by the concerns mentioned previously. I suspect that foreshortening information may be given more perceptual weight for reasons other than its theoretically determined reliability. For example, foreshortening information inheres locally in a texture (when isotropy applies), making it more informative in limited fields of view. It also is more informative for curved surfaces (see Section 6.2 below).

I should note, finally, that I have decomposed texture information in a very particular way. One need not decompose the information provided by texel shapes into foreshortening and scaling, as we have. The logic behind the decomposition is that the cues derive from separable effects of perspective distortion. A stronger case can be made by noting that one should distinguish scale-invariant shape information (which we have referred to as foreshortening) from size information (scaling), since the two rely on dissociable constraints on surface textures. In particular, application of an isotropy constraint is limited to the information pro-

vided by scale-invariant shape information. One could, however, derive different parameterizations of texture information in which foreshortening and scaling information are confounded. Malik and Rosenholtz [11], for example, characterize texture information implicitly as the affine relationships between neighboring regions in an image. Which representation best approximates human visual processing is a difficult challenge for future psychophysical work (see [32] for a first cut at testing what parameterization the visual system uses for estimating surface curvature from texture).

6.1.3. Field of view effects

The simulations in which we analyzed the effects of field of view have two major implications. First, to the extent that spatial integration of texture information is limited, attention should be focused on further parts of a surface to obtain reliable texture information about surface orientation, since it is there that the most reliable information about surface orientation is to be found. This observation is limited however to situations in which surfaces are constrained to be planar, or near planar. Somewhat surprisingly, particularly for a gradient cue like perspective scaling, increasing the field of view by including regions of the image in which the surface is closer to the observer (the bottom of images of ground planes) adds little to the informativeness of the texture information, except what is gained by obtaining more texture samples. Second, the relative reliability of different texture cues varies with field of view. Blake et al. [10] originally pointed this out in their analysis of foreshortening and density information. Not surprisingly similar observations pertain to scaling information. Blake et al. [30] have tested for an effect of field of view on the relative informativeness of texture cues and found no significant effects, with foreshortening being given the most weight by subjects under both small and large fields of view.

6.1.4. Isotropy

Throughout most of this paper, I have implicitly assumed isotropy in deriving estimators for surface orientation from texture and in our analysis of the reliability of foreshortening information. In the final set of simulations I relaxed this assumption by assuming that surface textures could be globally compressed by arbitrary amounts. The resulting ideal observer is appropriate for images of such anisotropic textures, however, is not ideal for isotropic textures. The results of the simulations showed it to be a very inefficient estimator of surface orientation for textures which are isotropic (4% efficiency for most surface slants). This testifies to the power of the isotropy constraint, when it is appropriate. Psychophysical data suggests that the human visual system applies an isotropy constraint to interpret planar surface orientation from texture [2,27,31], though, the strength of the constraint remains an open question.

The power of the isotropy constraint and the observation that not all textures are isotropic suggest that a worthwhile estimation strategy is to simultaneously estimate surface orientation and test the isotropy constraint. Applying a statistical estimation approach as used here provides a framework for testing assumptions such as isotropy [9,33]. An interesting question which arises is whether the human visual system can dynamically determine the applicability of a constraint like isotropy from image data.

6.2. Generalization to curved surfaces

The current paper has focused entirely on planar surfaces. One can, however, extend the results to some degree to curved surfaces. The ideal and generic observer formulation generalizes naturally to curved surfaces. One need simply replace the slant and tilt parameters in the model with appropriate shape parameters to characterize curved surfaces. The major limitation of this approach is that it requires a global parameterization of surface geometry. Since our major concern has been with analyzing the information content of texture, such a limitation is unavoidable and entirely appropriate. For computer vision applications, however, the limitation may be overly constraining and it is likely that ideal observer formulations like the one presented here will only be applicable in pure form to very constrained domains and will have to be integrated with more heuristic techniques in actual application.

Other authors, (e.g. [29]) have commented on the fact that foreshortening information is significantly more salient for curvature estimation than is scaling information. This derives, for isotropic textures, from the fact that foreshortening is a local cue, whereas scaling requires significant changes in depth to be informative. Psychophysical work has confirmed the relative importance of foreshortening information for the perception of curvature [29,32,34]. The salient portion of the current analysis, therefore, pertains to foreshortening information. Two aspects of the results have implications for curvature perception—the effect of surface slant on texture reliability and the importance of isotropy for the foreshortening cue.

Just as the orientation of planar surfaces becomes less ambiguous with increasing surface slant, so to should the curvature of surfaces. One often has the phenomenal percept in images of curved, textured surfaces that the curvature of those parts of the surface which are highly slanted away from the line of sight is much clearer than in other parts of the surface. For particular classes of surfaces, the observation suggests that certain viewing conditions will be more informative than others. For example, cylindrical surfaces oriented

out of the fronto-parallel plane should give rise to texture information which is much more reliable than cylindrical surfaces oriented in the fronto-parallel plane (as are typically used in psychophysical experiments).

The current analysis suggests a dramatic effect of violating isotropy on shape from texture perception. Some corroboration of this prediction has been obtained in psychophysical studies [34], though other results [32] suggest that the isotropy assumption may apply primarily to the determination of local surface tilt. Since the curvature in the stimuli used in the former study was entirely captured by spatial derivatives of slant (images of cylinders oriented in the fronto-parallel plane), the existing studies do not resolve the question of what role is played by isotropy in the interpretation of surface curvature from texture.

Appendix A

For the ideal observer simulations described in the text, I used spherical perspective as the model of perspective projection; however, the effects of perspective projection are easier to visualize using the more standard planar perspective. In this section, I will therefore derive the perspective map and its differential planar perspective. For ease of exposition, we will temporarily assume a fixed surface tilt in the vertical direction. A natural coordinate frame for surfaces slanted away from the viewer in a vertical direction is given by the basis set,

$$\{\vec{e}_1, \vec{e}_2\} = \{(1, 0, 0)^T, (0, \cos\sigma, \sin\sigma)^T\} \tag{43}$$

where the global slant σ is measured relative to the fronto-parallel (the x–y plane of the 3-D reference frame). Position in the surface plane is then specified by the 2-D coordinates, $\hat{U} = (u, v)^{\mathrm{T}}$, where positions in the plane are mapped to positions in 3-D world coordinates by $\hat{\mathrm{X}}_w = u\hat{\mathrm{e}}_1 + v\hat{\mathrm{e}}_2$.

Assuming that the surface is a distance d from the center of projection and that projection is onto a fronto-parallel projection surface at the same distance d from the center of projection, we have for the image coordinates of a point on the surface

$$\vec{X}_I(u, v) = \begin{pmatrix} x(u, v) \\ y(u, v) \end{pmatrix} = \begin{pmatrix} \dfrac{ud}{d + v\sin\sigma} \\ \dfrac{vd\cos\sigma}{d + v\sin\sigma} \end{pmatrix} \tag{44}$$

The differential of the projective map is given by the matrix

$$P = \begin{pmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{pmatrix} = \begin{pmatrix} \dfrac{d}{d + v\sin\sigma} & -\dfrac{ud\sin\sigma}{(d + v\sin\sigma)^2} \\ 0 & \dfrac{d^2\cos\sigma}{(d + v\sin\sigma)^2} \end{pmatrix} \tag{45}$$

which can be re-formulated in terms of projected position in the image as

$$P = \left(1 - \frac{y}{d}\tan\sigma\right)\begin{pmatrix} 1 & -\dfrac{x}{d}\sin\sigma \\ 0 & \cos\sigma - \dfrac{y}{d}\sin\sigma \end{pmatrix} \tag{46}$$

the equation used in the text.

Appendix B

I derive here the relationship between the second-order spatial moments of a surface texel and its projected image. A surface texel is a simple bounded region on the surface, R_{S}, which projects to an image texel defined as a bounded region in the image, R_{I}. Points in R_{I} are related to points in R_{S} by the equation

$$\vec{X} = f(\vec{U}) \tag{47}$$

where $\vec{X}$ is position in the image plane and $\vec{U}$ is position on the surface. One can approximate the mapping of points within a surface texel to the image by a Taylor series expansion of f() around the center of mass of the texel. Discarding the higher-order terms, we obtain the affine approximation,

$$\vec{X} = f(\vec{U}_0) + P(\vec{U} - \vec{U}_0) + R(O^2) \tag{48}$$

where $\vec{U}_0$ is the center of mass of the surface texel and P is the Jacobian of f(), computed at $\vec{U}_0$ (see Appendix A for a derivation of P).

The second-order spatial moments (moments of inertia) of an image texel are given by the matrix

$$M_{\mathrm{I}} = \frac{1}{A_{\mathrm{I}}}\int\int_{R_{\mathrm{I}}} (\vec{X} - \vec{X}_0)(\vec{X} - \vec{X}_0)^T \,\mathrm{d}A_{\mathrm{I}} \tag{49}$$

$$M_{\mathrm{I}} = \frac{1}{A_{\mathrm{I}}}\begin{pmatrix} \displaystyle\int\int_{R_{\mathrm{I}}} (x - x_0)^2 \,\mathrm{d}A_{\mathrm{I}} & \displaystyle\int\int_{R_{\mathrm{I}}} (x - x_0)(y - y_0) \,\mathrm{d}A_{\mathrm{I}} \\ \displaystyle\int\int_{R_{\mathrm{I}}} (x - x_0)(y - y_0) \,\mathrm{d}A_{\mathrm{I}} & \displaystyle\int\int_{R_{\mathrm{I}}} (y - y_0)^2 \,\mathrm{d}A_{\mathrm{I}} \end{pmatrix} \tag{50}$$

where A_I is the area of the image texel, $\vec{X}_0$ is the center of mass of the image texel and dA_I is a differential area element in the image. In order to derive the relationship between M_I and M_S (the moments of the surface texel), we first need to derive terms for the area and the center of mass of an image texel.

The area of an image texel is given by

$$A_I = \int\int_{R_I} dA_I = \int\int_{(x,y)\in R_I} dx\, dy \tag{51}$$

Transforming variables from image coordinates to surface coordinates, we obtain for the area

$$A_I = \int\int_{(u,v)\in R_S} \det(P(u,v))\, du\, dv \tag{52}$$

where, as before, P is the Jacobian of the coordinate transform. Since we approximate the projection of points within a texel to be an affine transformation, P is constant and we obtain

$$A_I = \det(P)\, A_S \tag{53}$$

where A_S is the area of the surface texel.

The center of mass of an image texel is given by

$$\vec{X}_0 = \frac{1}{A_I}\int\int_{(x,y)\in R_I} \vec{X}\, dx\, dy \tag{54}$$

Again changing variables to surface coordinates and substituting $\det(P)A_S$ for A_I, we obtain

$$\vec{X}_0 = \frac{1}{\det(P)\, A_S}\int\int_{(u,v)\in R_S} [f(\vec{U}_0) + P(\vec{U}-\vec{U}_0)]\det(P) \times du\, dv \tag{55}$$

$$\vec{X}_0 = \frac{1}{A_S}\left[\int\int_{(u,v)\in R_S} f(\vec{U}_0)\, du\, dv + P\left(\int\int_{(u,v)\in R_S} \vec{U}\, du\, dv - \int\int_{(u,v)\in R_S} \vec{U}_0 du\, dv\right)\right] \tag{56}$$

By the definition of the center of mass, the second term in the expression goes to zero and we are left with

$$\vec{X}_0 = \frac{1}{A_S}\int\int_{(u,v)\in R_S} f(\vec{U}_0)\, du\, dv \tag{57}$$

$$\vec{X}_0 = f(\vec{U}_0) \tag{58}$$

that is, the center of mass of the image texel is simply the projection of the center of mass of the surface texel.

It is now straightforward to derive the relationship between the moments of an image texel and its corresponding surface texel. A change of variables in Eq. (50) from image to surface coordinates gives

$$M_I = \frac{1}{A_I}\int\int_{(x,y)\in R_I} (\vec{X}-\vec{X}_0)(\vec{X}-\vec{X}_0)^T\, dx\, dy \tag{59}$$

$$M_I = \frac{1}{\det(P)\, A_S}\int\int_{(u,v)\in R_S} P(\vec{U}-\vec{U}_0)(P(\vec{U}-\vec{U}_0))^T \times \det(P)\, du\, dv \tag{60}$$

$$M_I = P\left[\frac{1}{A_S}\int\int_{(u,v)\in R_S} (\vec{U}-\vec{U}_0))(\vec{U}-\vec{U}_0)^T\, du\, dv\right]P^T \tag{61}$$

The integral term is simply the expression for the second order moments of the surface texel, so we have, finally,

$$M_I = PM_SP^T \tag{62}$$

References

[1] Gibson JJ. The Perception of Visual Surfaces. Am J Psychol 1950;63:367–84.

[2] Frisby JP, Buckley D. Experiments on stereo and texture cue combination in human vision using quasi-natural viewing. In: GaN Orban HH editor. Artificial and Biological Visual Systems. Berlin:Springer-Verlag, 1992.

[3] Johnston E, Cumming B, Parker A. Integration of depth modules: stereopsis and texture. Vis Res 1993;33(5/6):813–26.

[4] Young M, Landy M, Maloney L. A perturbation analysis of depth perception from combinations of texture and motion cues. Vis Res 1993;33(18):2685–96.

[5] Knill DC. Discriminating surface slant from texture: comparing human and ideal observers. Vis Res 1998;38:1683–1711.

[6] Witkin AP. Recovering Surface Shape and Orientation from Texture. Artif Intell 1981;17(1):17–45.

[7] Davis LS, Janos L, Dunn SM. Efficient recovery of shape from texture. IEEE PAMI 1993;5(5):485–92.

[8] Kanatani L, Chou T. Shape from texture: general principles. Artificial Intelligence 1989;38:1–48.

[9] Blake A, Marinos C. Shape from texture: estimation, isotropy and moments. Artif Intell 1990;45:323–80.

[10] Blake A, Bulthoff HH, Sheinberg A. Shape from texture: ideal observers and human psychophysics. Vis Res 1993;33(12):1723–37.

[11] Malik J, Rosenholtz R. Recovering surface curvature and orientation from texture distortion: a least squares algorithm and sensitivity analysis. Proc. 3rd European Conf. on Computer Vision, Volume 800 of Lecture Notes in Computer Science. Berlin: Springer-Verlag, 1995:353–364.

[12] Garding J. Shape from texture for smooth curved surfaces in pespective projection. J Math Imag Vis 1992;2(4):327–50.

[13] Garding J. Surface orientation and curvature from differential texture distortion, in Proc. 5th International Conference on Computer Vision, Cambridge, MA, 1995:733–739.

[14] Knill DC. Estimating illuminant direction and degree of surface relief. J Opt Soc Am 1990;7A(4):759–75.

[15] Stevens KA. The information content of texture gradients. Biol Cybernet 1981;42:95–105.

[16] Bajcsy R, Lieberman L. Texture gradients as a depth cue. Comput Graph Imag Process 1976;5:52–67.

[17] Super B, Bovik AC. Shape from texture using local spectral moments. IEEE PA MI 1995;17:333–43.

[18] Sakai K, Finkel LH. Characterization of the spatial frequency spectrum in perception of shape-from-texture. J Opt Soc Am 1995;12A:1208–24.

[19] Lindeburg T, Garding J. Shape from texture from a multi-scale perspective, in Proc. 4th Int. Conf. Comput Vis 1993:683–691.

[20] Freeman WF. The generic viewpoint assumption in a framework for visual perception. Nature 1994;368:542–5.

[21] Brown LG, Shvayster H. Surface orientation from projective foreshortening of isotropic texture autocorrelation. IEEE Tran. Pattern Anal Machine Intell 1990;12(6):584–8.

[22] Shannon CE, Weaver W. The mathematical theory of communication, University of Illinois Press, Chicago, 1963.
[23] Rosenholtz R, Malik J. Shape from texture: Isotropy or homogeneity (or both)? Vis Res 1997;37:2283–2294.
[24] Rao C. Linear statistical inference and its applications. New York: Wiley, 1973.
[25] Fortune SJ. A sweepline algorithm for Voronoi diagrams. Algorithmica 1987;2(15):3–174.
[26] Arnold L. Stochastic differential equations: theory and applications, Krieger Press, Malabar, FL, 1992.
[27] Rosenholtz R, Malik J. An ideal observer model for shape from texture. Invest Opthalmol Vis Sci 1994;35(Supplemental issue 4):1668.
[28] Braunstein ML, Payne JW. Perspective and form ratio as determinants of relative slant judgments. J Exp Psychol 1969;81(3):584–90.
[29] Cutting J. Three gradients and the perception of flat and curved surfaces. J Exp Psychol: general 1984;113(2):198–216.
[30] Buckley D, Frisby J, Blake A. Does the human visual system implement an ideal observer theory of slant from texture? Vis Res 1996;36(8):1163–76.
[31] Knill D. Ideal observer perturbation analysis reveals human strategies for inferring surface orientation from texture, Vis Res (in press).
[32] Todd J, Akerstrom R. Perception of three-dimensional form from patterns of optical texture, Journal of Experimental Psychology: Human Perception and Performance 1987;13:242–255.
[33] Yuille A, Bulthoff HH. Bayesian decision theory and psychophysics. In: Knill DC, Richards W. editors. Perception as Bayesian Inference. Cambridge Univ. Press, New York, 1996.
[34] Cumming BG, Johnston EB, Parker AJ. Effects of different texture cues on curved surfaces viewed stereoscopically. Vis Res 1993;33(5–6):827–38.